Four Essays on Robustification of Portfolio Models

Dissertation

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaften (Dr. rer. pol.)
am Fachbereich Wirtschaftswissenschaften
der Universität Konstanz

vorgelegt von:
Peter Schanbacher
St. Gebhardstrasse 26
78467 Konstanz

Tag der mündlichen Prüfung: 13.12.2013

1. Referent: Prof. Dr. Winfried Pohlmeier
2. Referent: Prof. Dr. Axel Kind

To:

my wife Conny and

my parents Ute & Roland

Acknowledgement

This doctoral dissertation has been written at the Department of Economics at the University of Konstanz. I am indebted to my supervisor Prof. Dr. Winfried Pohlmeier for his continuous support of my work and his professional advice and guidance.

I wish to express my thanks to Prof. Dr. Axel Kind who agreed to serve as the second supervisor of my dissertation.

Special thanks go to my colleagues Ruben Seiberlich, Fabian Krüger, Christoph Frey, Frieder Mokinski, Hao Liu and Lidan Li who always offered me friendship and advise. I am particularly thankful to my colleagues and friends Steffen Schaarschmidt and Matthias Draheim for uncountable hours of enjoyable and successful joint work and entertaining coffee breaks.

Obviously without my parents this thesis (as well as myself) would not have been developed. I owe them my deepest thanks for their kindness, their continuous support and their helpful advice in every decision I have made.

Last but not least my deepest thanks go to my wife Conny for her love, her patience, her support and her encouragement.

Contents

List of Tables

List of Figures

Summary

The financial crisis has shown that quantitative asset allocation models and risk management models have not been sufficiently understood. Since the seminal paper of Markowitz (1952), the academic portfolio management literature has relied on sample estimates of return moments, which have led to extreme and unstable portfolio weights. In the 1990s, Britten-Jones (1999) and others demonstrated the poor empirical performance of the standard Markowitz model. As particularly the mean is difficult to estimate, literature often relies on the minimum variance portfolio (for a sensitivity analysis to changes to the mean see Best and Grauer, 1991a). Still, the estimation error in the covariance matrix is large (Chan et al., 1999). To reduce the estimation error in the covariance matrix, Ledoit and Wolf (2004) suggest a shrinkage approach.
Several regularization procedures have been proposed. Jagannathan and Ma (2003) advise short-selling restrictions. The approach is generalized by Brodie et al. (2009) who penalize short positions. They introduce a Lasso penalty on the norm of the portfolio weights. Stabilizing portfolio weights by shrinking them directly towards some predefined target is studied by Frahm and Memmel (2010). Despite numerous efforts, DeMiguel et al. (2009b) show that naive portfolio strategies such as the equally weighted portfolio are difficult to outperform. Clearly, it is not yet fully understood under which circumstances portfolio optimization fails. This thesis advocates and contributes to the development of alternatives and extensions to current portfolio optimization procedures.
One possibility to reduce estimation risk is the combination of different asset allocation models. Model combination in a forecasting context is analyzed in Chapter 1, while the transfer to asset allocation is presented in Chapter 2. The combined model turns out to have a stable performance and does not (strongly) suffer from misspecification of the individual models. Other possibilities are alternative asset allocation strategies or improvements upon current portfolio optimization procedures. Chapter 3 advocates a portfolio strategy, called Minimax, which does not suffer

from estimation risk in the returns' moments. We show that the Minimax can be considered as an alternative to Markowitz portfolio optimization. An extension to standard portfolio optimization is given in Chapter 4. We introduce a penalty term for the norm of portfolio weights to prevent extreme asset allocations. The penalty is shown to improve the performance of standard asset allocation strategies.

Chapter 1 is a joint paper with Fabian Krüger and analyzes the performance of model combination. The performance of a probabilistic forecast model is commonly measured by strictly proper scoring rules (Gneiting and Raftery, 2007). It has been shown that some popular scoring rules are concave functions of the forecast. By Jensen's inequality it holds that the average score is necessarily smaller (i.e. worse) than the score of the average forecast. This feature is often related to the good empirical performance of forecast combination, compared to the individual forecasts. The success of forecast combination is partly a consequence of the forecast evaluation methodology. We generalize the literature by showing that (smooth and strictly proper) scoring rules cannot be entirely convex, may be entirely concave, and are at least locally concave around the true probability. The finding implies that if forecast predictions are sufficiently close to the true probability, the performance of the average forecast is at least as good as the average performance of the individual models. Concavity depends on the true probability, which is not known in practice. For a given set of forecasts, we suggest to derive a range of true probabilities for which concavity holds. As an example, we analyze the prediction of US recessions based on a Probit and a survey of experts. We find that the range of probabilities under which the (spherical) score is concave is typically much larger than the interval defined by the two forecasts. Further, we find that the ex-post better model and the combined model significantly outperform the ex-post worse model. However, the ex-post better model and the combined model are statistically indistinguishable. We conclude that model combination is rewarding for most scenarios and scoring rules.

Chapter 2 transfers the insights gained in Chapter 1 to portfolio optimization. Under a (strictly) concave performance function, the average model is better than the average performance of the individual models. We find that many performance measures used in asset allocation are concave, while others are concave under certain conditions. So far, the literature was hunting for one single "true" asset allocation model. We run a large empirical study to analyze the average model of several well-known asset allocation models. We use five models of three model classes applied to six different data sets and evaluate them by five different performance measures. We find that no single model constantly outperforms the others. The ranking of the models depends on the performance measure as well as on the chosen data set. Even for a certain performance measure and a certain data set, the ranking of the models strongly changes over time. The finding confirms that in calm periods sophisticated models outperform naive models; in rough periods data-ignorant models outperform sophisticated models. In the situation of changing model ranking and concave performance measures, the average model has to perform well by definition. The theoretical conclusion can be affirmed by our empirical study. The average model performs almost as good as the ex-post best model.

Chapter 3 is a joint paper with Steffen Schaarschmidt proposing a portfolio strategy called Minimax. The strategy deviates from classical risk measures and defines risk in terms of the worst case scenario. Common symmetric risk measures have undesirable properties. First, large positive returns should not be considered as risk. Second, rare extreme losses might get too little attention. In view of the recent economic turmoil, investors may prefer our conservative risk measure. Additionally, our approach circumvents the estimation of unstable means as well as the estimation of a large covariance matrix. A typical investor might be a corporation, a pension fund, or a bank which has to implement daily risk management. The target is to minimize daily investment losses. A second type of investor is an investor who is facing mark-to-market accounting. He aims to minimize the margin calls as the portfolio falls short of a certain level. The Minimax strategy is a "pessimistic" trading strategy, as it chooses the portfolio weights such that the portfolio payoff is *max*imized for the *min*imum outcome. Our contribution is to show that the Minimax portfolio is implementable for a multi asset investor. In our empirical study, we use US stock, bond, real estate and commodity indexes to construct portfolios with yearly holding periods. We compare the performance of the Minimax portfolio to the performance of the asset allocation of a typical pension fund. Additionally, we use alternative asset allocation strategies, such as the equally weighted portfolio, the Minimum Variance portfolio and the (short-selling restricted) Mean Variance portfolio. We find that the Minimax portfolio performs well compared to the benchmarks considered.

Chapter 4 is a joint paper with Prof. Dr. Pohlmeier considering a recent regulariza-

tion approach for portfolio weights. To stabilize portfolio weights, Jagannathan and Ma (2003) find that the no-shortsale constraint works well in portfolio optimization. Still, the approach is too restrictive, as moderate short positions can enhance the performance of the portfolio (Fan, 2010). Brodie et al. (2009) introduce a L_1 norm restriction for the portfolio weights. In the statistics literature, the type of restriction was introduced as the "Lasso" by Tibshirani (1996). The restriction can be interpreted as a penalization of the portfolio's short positions. Naturally, the question arises how to determine the optimal penalty level for short-selling. Our contribution is to introduce a rule-of-thumb for the penalty level. The resulting "Lasso portfolio" is easy to implement and asymptotically optimal. It performs well in a simulation study. In an empirical study, we find that the Lasso portfolio outperforms the no short-sale Mean Variance portfolio, the unrestricted Mean Variance portfolio as well as various alternative strategies proposed by literature.

Zusammenfassung

Die Finanzkrise hat gezeigt, dass quantitative Modelle zur Portfoliostrukturierung sowie auch Risikomodelle noch starke Schwächen besitzen. Nach der wegweisenden Arbeit von Markowitz (1952) vertraute die akademische Portfolio Management Literatur zu lange auf die Stichprobenschätzer der Renditenmomente. Diese führten zu extremen und unstabilen Portfoliogewichten. In den 90ern analysierten Britten-Jones (1999) und weitere Forscher die schwachen empirischen Ergebnisse des Markowitz Modells. Da insbesondere der Erwartungswert schwer zu schätzen ist, wird in der Literatur oft das Minimum-Varianz Portfolio verwendet (für eine Sensitivitätsanalyse siehe Best und Grauer, 1991a). Aber auch der Schätzfehler der Kovarianzmatrix ist groß (Chan et al., 1999). Um den Schätzfehler der Kovarianzmatrix zu reduzieren, schrumpfen Ledoit und Wolf (2004) die geschätzte Kovarianzmatrix zu einer Zielmatrix.

Einige alternative Regularisierungsmethoden wurden vorgeschlagen. Jagannathan und Ma (2003) empfehlen Leerverkäufe zu vermeiden. Diesen Ansatz verallgemeinern Brodie et al. (2009), indem sie Leerverkäufe bestrafen. Sie führen eine Lasso-Bestrafung für die Norm der Portfoliogewichte ein. Eine Stabilisierung der Portfoliogewichte durch die Verschiebung zu vordefinierten Gewichten untersuchen Frahm und Memmel (2010). Trotz aller Anstrengungen zeigen DeMiguel et al. (2009b), dass einfache Strategien, wie das gleichverteilte Portfolio, schwer zu schlagen sind. Offensichtlich ist noch nicht voll verstanden warum und unter welchen Umständen Portfolio-Optimierung versagt. Die vorliegende Dissertation trägt zur Entwicklung von Alternativen und Erweiterungen der bisherigen Portfolio-Optimierung bei. Eine Möglichkeit das Schätzrisiko zu reduzieren, ist die Kombination von verschiedenen Anlagemodellen. Modellkombinationen werden in Kapital 1 untersucht und in Kapitel 2 auf Anlagestrategien übertragen. Das kombinierte Modell zeigt ein stabileres Ergebnis und leidet nicht (stark) an den Schätzfehlern der individuellen Modelle. Weitere Möglichkeiten sind alternative Strategien, oder die Verbesserung bisheriger Methoden. Kapitel 3 empfiehlt eine

Portfoliostrategie, genannt Minimax, die nicht vom Schätzrisiko der Momente der Renditen abhängig ist. Wir zeigen, dass die Minimax-Optimierung als Alternative zur Markowitz Optimierung verwendet werden kann. Eine Erweiterung zur Portfolio-Optimierung ist in Kapitel 4 gegeben. Wir führen einen Bestrafungsterm für die Norm von Portfoliogewichten ein um extreme Anlagen zu vermeiden.
Kapitel 1 ist eine gemeinsame Arbeit mit Fabian Krüger und untersucht die Ergebnisse von Modellkombinationen. Die Leistung einer Wahrscheinlichkeitsvorhersage wird im Allgemeinen mit strikt ordentlichen Bewertungsregeln gemessen (Gneiting und Raftery, 2007). Einige bekannte Bewertungsregeln sind konkave Funktionen der Vorhersage. Die Jensenschen Ungleichung impliziert, dass die durchschnittliche Bewertung notwendigerweise kleiner (d.h. schlechter) als die Bewertung der durchschnittlichen Vorhersage ist. Diese Eigenschaft wird oft mit der empirisch guten Leistung von Vorhersagekombinationen in Verbindung gebracht. Der Erfolg der Kombination von unterschiedlichen Vorhersagen ist teilweise eine Konsequenz der Auswertungsmethode. Wir verallgemeinern die Literatur, indem wir zeigen, dass (glatte und strikt korrekte) Bewertungsregeln nicht vollständig konvex sein können. Jedoch können sie vollständig konkav sein und sind zumindest lokal konkav um den wahren Wahrscheinlichkeitswert. Das Ergebnis impliziert, dass die Leistung der durchschnittlichen Vorhersage besser ist als die durchschnittliche Leistung der einzelnen Vorhersagen, falls die Vorhersagen genau genug sind. Konkavität hängt von der wahren Wahrscheinlichkeit ab, welche in der Realität unbekannt ist. Für gegebene Vorhersagen empfehlen wir ein Intervall von wahren Wahrscheinlichkeiten anzugeben, für welche Konkavität zutrifft. Als Beispiel analysieren wir die Vorhersage von Rezessionen in den USA basierend auf einem Probitmodell und einer Expertenumfrage. Das Intervall, für welches die (kugelförmige) Bewertungsregel konkav ist, ist typischerweise weit grösser als das Intervall beider Vorhersagen. Des Weiteren stellen wir fest, dass das rückwirkend bessere Modell sowie die Kombination das schlechtere Modell signifikant schlagen. Jedoch sind die Kombination und das bessere Modell statistisch nicht zu unterscheiden. Wir folgern, dass die Kombination von Modellen für die meisten Fälle und Bewertungsregeln lohnenswert ist.

Kapitel 2 überträgt die in Kapitel 1 gewonnenen Einblicke auf die Portfolio-Optimierung. Unter einem (strikt) konkaven Leistungsmaß besitzt das durchschnittliche Modell eine überdurchschnittliche Leistung. Wir zeigen, dass viele

Leistungsmaße in der Portfolio-Optimierung konkav sind, während andere unter bestimmten Bedingungen konkav sind. Bisher war die Literatur auf der Suche nach einem einzelnen, "wahren" Anlagemodell. Wir führen eine große empirische Studie durch, um das durchschnittliche Modell einiger bekannter Anlagestrategien zu analysieren. Wir verwenden fünf Modelle aus drei Modellkategorien, angewandt auf sechs Datensätze und ausgewertet mittels fünf unterschiedlicher Leistungsmaße. Wir beobachten, dass kein Modell durchweg ein anderes schlägt. Das beste Modell hängt sowohl vom Leistungsmaß als auch vom gewählten Datensatz ab. Selbst für ein bestimmtes Leistungsmaß und einen festen Datensatz verändert sich im Laufe der Zeit die Reihenfolge der Modelle stark. Die Beobachtung bestätigt, dass in ruhigen Zeiten hochentwickelte Modelle einfache Modelle schlagen, während in unruhigen Zeiten einfache Modelle die hochentwickelten Modelle schlagen. Angesichts der sich ändernden Reihenfolge der Modelle und der konkaven Leistungsmaße muss das durchschnittliche Modell per Definition eine gute Leistung besitzen. Diese theoretische Überlegung wird durch unsere empirische Studie bestätigt. Das durchschnittliche Modell zeigt eine nahezu gleich gute Leistung wie das rückwirkend beste Modell.

Kapitel 3 ist eine gemeinsame Arbeit mit Steffen Schaarschmidt und schlägt eine neue Portfolio Strategie, genannt Minimax, vor. Mit dieser Strategie weichen wir von klassischen Risikomaßen ab und definieren Risiko mittels des ungünstigsten Falles. Verbreitete symmetrische Risikomaße haben unerwünschte Eigenschaften. Erstens sollten große positive Renditen nicht als Risiko gewertet werden. Zweitens erhalten seltene, extreme Verluste eventuell zu wenig Beachtung. In den aktuellen wirtschaftlichen Turbulenzen könnte ein Investor unser konservatives Risikomaß bevorzugen. Zusätzlich vermeidet unser Ansatz die Schätzung von unstabilen Erwartungswerten und großen Kovarianzmatrizen. Typische Investoren sind Unternehmenen, Pensionsfonds oder Banken, die täglichem Risikomanagement ausgesetzt sind. Das Ziel der Strategie ist tägliche Verluste zu minimieren. Der Ansatz ist auch für Investoren interessant, die dem Marktbewertungsansatz ausgesetzt sind. Sie versuchen die Nachschussforderung zu minimieren, die fällig wird, wenn das Portfolio unter ein gewisses Level fällt. Die Minimax Strategie ist eine "pessimistische" Handelsstrategie, da die Portfoliogewichte so gewählt werden, dass im Falle des minimalen Ausgangs die Portfolio Auszahlung maximiert wird. Wir zeigen, dass das Minimax-Portfolio für einen Investor mit mehreren Anlagen implementierbar ist.

Für unsere empirischen Studie verwenden wir US Aktien-, Anleihen-, Immobilien- und Rohstoffindizes um Portfolien mit jährlicher Halteperiode zu konstruieren. Wir vergleichen die Leistung des Minimax-Portfolios mit der Leistung der Anlagestrategie eines typischen Pension Fonds. Zusätzlich verwenden wir alternative Anlagestrategien, wie das gleichgewichtete Portfolio, das Minimum-Varianz Portfolio und das Erwartungswert-Varianz Portfolio (ohne Leerverkäufe). Wir stellen fest, dass sich das Minimax Portfolio im Vergleich zu den Alternativportfolios gut verhält.

Kapitel 4 ist eine gemeinsame Arbeit mit Prof. Dr. Pohlmeier, welches einen aktuellen Regularisierungsansatz für Portfoliogewichte betrachtet. Um Portfoliogewichte zu stabilisieren zeigen Jagannathan und Ma (2003), dass ein Verbot von Leerverkäufen sich positiv auf die Portfolioleistung auswirkt. Dieser Ansatz ist zu restriktiv, da geringe Leerverkäufe die Leistung des Portfolios verbessern können (Fan, 2010). Brodie et al. (2009) führen eine L_1-Norm Beschränkung für Portfoliogewichte ein. Diese Art von Beschränkung wurde ursprünglich von Tibshirani (1996) unter dem Namen "Lasso" eingeführt, und kann in diesem Zusammenhang als Bestrafung von Leerverkäufen interpretiert werden. Die Frage nach dem optimalen Bestrafungslevel für Leerverkäufe beantworten wir durch eine einfache Regel. Das resultierende "Lasso-Portfolio" ist leicht zu implementieren und asymptotisch optimal. In der Simulationsstudie zeigt es eine gute Leistung. In der empirischen Studie schlägt das Lasso-Portfolio das Erwartungswert-Varianz Portfolio sowohl mit als auch ohne Leerverkäufe.

Chapter 1

Concave Scoring Rules

1.1 Introduction

Probabilistic forecasts are used in an increasing number of applications in economics. *Strictly proper scoring rules* (Gneiting and Raftery, 2007) are loss functions for probabilistic forecasts. In expectation, these scoring rules are maximized by the true probability model. Thus, they are a useful tool for comparing alternative probabilistic forecasts, and complement diagnostic approaches based on the calibration and sharpness of these forecasts (Diebold et al., 1998; Gneiting et al., 2007).

Staël von Holstein (1970, p. 69) shows that some of the most popular scoring rules, such as the logarithmic score (Good, 1952), are concave functions of the forecast. Hence by Jensen's inequality, the average score of several forecasts is necessarily smaller than the score of the average (combined) forecast.[1] Lichtendahl et al. (2012) relate this feature to the good empirical performance of combined probabilistic forecasts, as compared to the individual forecasts which enter the combination. The intriguing implication of these papers is that the well documented success of forecast combinations (sometimes called the "wisdom of the crowd") is partly a tautological consequence of the evaluation methodology.

The finding cited above refers to specific (strictly proper) scoring rules, of which there are infinitely many. For example, the results by Shuford et al. (1966), Hendrickson and Buehler (1971) and Schervish (1989) imply general recipes to construct such rules. This raises the question whether more general statements about the concavity of strictly proper scoring rules can be made. This question has not been addressed so far; the present paper aims to fill this gap.

Focusing on a simple binary prediction problem, we derive some simple but novel results about strictly proper scoring rules. First, we show that they cannot be entirely convex.[2] They may be, but need not be, entirely concave. For example, the results by Staël von Holstein imply that the logarithmic and Brier (1950) scores are entirely concave while the spherical score (Toda, 1963) is not. Second, we show that any (smooth) strictly proper scoring rule is locally concave around the true probability.

[1] This finding is similar to results from the literature on combining point forecasts (see e.g. Larrick and Soll, 2006; Manski, 2010, and the references therein).
[2] A precise definition of the word "entirely" is provided below.

Our results imply that if the scoring rule is not entirely concave, concavity depends on the (unknown) true probability of an event. This suggests to derive a range of true probabilities for which concavity holds, given a certain set of forecasts. As a simple heuristic, a researcher could then decide to combine if he or she is willing to assume that the true probability lies within this range. We apply this idea to the task of predicting US recessions. The first forecast is from a survey of economists, while the second is based on a simple Probit regression based on the spread between interest rates at different maturities. Our analysis focuses on the spherical score. In our example, we find that the range of probabilities under which the spherical score is concave typically exceeds the interval defined by the two alternative forecasts. That is, concavity holds unless both forecasts are severely misspecified.

The remainder of this paper is organized as follows. Section 1.2 analyzes the concavity of strictly proper scoring rules. Section 1.3 discusses the implications of concavity for the performance of linear forecast combinations. Section 1.4 contains the empirical application described above. Section 1.5 provides a concluding discussion.

1.2 Concavity

We focus on a binary random variable $Y \in \{0,1\}$, and consider probabilistic prediction of this random variable. The set of admissible probability distributions is indexed by a scalar $p \in [0,1]$ which denotes the predicted probability that $Y = 1$.

Scoring rules allow to evaluate the quality of the prediction p. We take scoring rules to be positively oriented (the greater the score, the better the prediction). The score associated with a binary outcome Y and a predicted probability p is given by

$$\tilde{S}(Y,p) = Y g_1(p) + (1-Y)\ g_0(p); \tag{1.1}$$

since the outcome Y is stochastic, the score $\tilde{S}(Y,p)$ is a random variable. The two functions $g_y(p),\ y \in \{0,1\}$, indicate the score of a prediction p if an outcome $Y = y$ realizes. Denote by $q \in [0,1]$ the true probability that $Y = 1$. Then, the expected score of the prediction p is given by

$$E\left(\tilde{S}(Y,p)\right) = qg_1(p) + (1-q)g_0(p) =: S(q,p). \tag{1.2}$$

Clearly, $S(q,p)$ is not a random variable but a deterministic function of q and p. Throughout this paper, we impose the following regularity conditions on $S(q,p)$.

Assumption 1.2.1. *a) $S(q,p)$ is real-valued and finite, except possibly that $S(0,1) = -\infty$ and/or $S(1,0) = -\infty$.*

b) $S(q,p)$ is twice continuously differentiable with respect to p, for all $q,p \in [0,1]$.

The first condition is standard (c.f. Gneiting and Raftery, 2007, Definition 1). The second condition is stricter than required for some of our results, but imposing it simplifies our presentation. This does not seem to come at a significant cost since most of the strictly proper rules considered in the literature satisfy the condition (see examples below).

As mentioned earlier, propriety is an important feature of every "reasonable" scoring rule. It ensures that S sets incentives for truth-telling on the part of forecasters.

Definition 1. *A scoring rule S is called proper if $S(q,p) \leq S(q,q)$ for all $p,q \in [0,1]$. It is called strictly proper if the inequality holds strictly for all $p \neq q$.*

A scoring rule is a function which assigns a payoff to every prediction p. The aim of this paper is to analyze certain concavity properties of scoring rules with respect to p. However, it is not immediately clear how to define "concavity of a scoring rule": A scoring rule S can be characterized by both the score function $\tilde{S}(Y,p)$ and the expected score function $S(q,p)$. Furthermore, concavity of these functions may depend on the true probability q and/or the prediction p.

We resolve the first issue by focusing on the *expected* score function $S(q,p)$. Unlike the score function $\tilde{S}(Y,p)$, $S(q,p)$ provides an ex ante measure for the performance of p. That is, the performance measure is available before the realization $Y = y$ is known. Our notion of concavity is thus slightly different from that of Staël von Holstein (1970, p. 69 and Appendix 3), who defines concavity in terms of the functions $g_1(p)$ and $g_0(p)$ in Equation (1.1) above. We return to this aspect below. To resolve the second issue, we need a definition which is flexible enough to allow for the possibility that concavity holds only for certain parameter ranges of p or q. These concerns motivate the following definition.

Definition 2. *Fix a certain true probability* $q \in [0,1]$. *Then, a scoring rule* S *is called* <u>*concave*</u> *on a set* $\mathcal{D} \subseteq [0,1]$ *if for all* $p_1 \neq p_2 \in \mathcal{D}$ *and* $\lambda \in (0,1)$, *it holds that*

$$S\left(q, \lambda p_1 + (1-\lambda)p_2\right) \geq \lambda S(q,p_1) + (1-\lambda)S(q,p_2).$$

S is called strictly concave if the statement holds for "$\geq$" replaced by "$>$". S is called convex if the statement holds for "$\geq$" replaced by "$\leq$".

If $\mathcal{D}$ is an interval, then a necessary and sufficient condition for concavity is that $\frac{\partial^2 S(q,p)}{\partial p^2} \leq 0$ for all $p \in \mathcal{D}$. For general $\mathcal{D}$, there is no obvious relation between concavity and the second derivative of the expected score function. We return to this issue in Section 1.4 below.

Requiring Definition 2 to hold for any q leads to the following stronger notion of concavity.

Definition 3. *A scoring rule* S *is called* <u>*entirely concave*</u> *on a set* $\mathcal{D} \subseteq [0,1]$ *if* S *is concave on* $\mathcal{D}$ *for every* $q \in [0,1]$.

Importantly, note that our definition of "entirely concave" coincides with the definition of "concave" of Staël von Holstein (1970) on page 69.[3] We next state our first result.

Proposition 1.2.1. *Consider a set* $\mathcal{D} \subseteq [0,1]$ *which has at least two distinct elements. Then, there is no strictly proper scoring rule* S *which is entirely convex on* $\mathcal{D}$.

Proof. See appendix. □

The proposition implies that the (potential) convexity of S can only hold for a restricted range of q. That is, even for a "small" set $\mathcal{D}$ (e.g. a set containing only two points), there is always a value of q such that S is not convex on $\mathcal{D}$.

It is important to distinguish Proposition 1.2.1 from Theorem 1 of Gneiting and Raftery (2007). These authors show that, for any strictly proper scoring rule, the

[3] He defines a scoring rule to be concave if the two functions $g_1(p)$ and $g_0(p)$ in Equation (1.1) are concave in p. If this is the case, then the expected score $q\ g_1(p) + (1-q)\ g_0(p)$ is again concave in p for any $q \in [0,1]$, and the scoring rule is entirely concave. Conversely, for the scoring rule to be entirely concave, both $g_1(p)$ and $g_0(p)$ must be concave.

entropy function $S(q, q)$ is convex in q. In contrast to the entropy function, however, we consider the expected score function $S(q, p)$, for a fixed true distribution q and varying forecasts p.

Proposition 1.2.1 points to an interesting asymmetry: While no strictly proper scoring rule can be entirely convex on $\mathcal{D} \subseteq [0, 1]$, some well known (and strictly proper) scoring rules are entirely strictly concave on the full unit interval. For example,[4] the logarithmic score given by

$$g_y(p) = \begin{cases} \ln(p) & y = 1 \\ \ln(1-p) & y = 0 \end{cases}$$

and the Brier score given by

$$g_y(p) = \begin{cases} -(1-p)^2 & y = 1 \\ -p^2 & y = 0 \end{cases}$$

are entirely strictly concave on $[0, 1]$.

However, Proposition 1.2.1 does not imply that any strictly proper scoring rule must be entirely concave. For example, the spherical score given by

$$g_y(p) = \begin{cases} \frac{p}{\sqrt{p^2+(1-p)^2}} & y = 1 \\ \frac{1-p}{\sqrt{p^2+(1-p)^2}} & y = 0 \end{cases}$$

is strictly proper but not entirely concave on $[0, 1]$.

In the light of the last example, a natural follow-up question is: Is some weaker notion of concavity necessary for strict propriety? Our following result shows that this is indeed the case.

Proposition 1.2.2. *Fix a true probability $0 < q < 1$, and consider a strictly proper scoring rule S. Then, there is an $\varepsilon > 0$ such that S is concave on the interval $\mathcal{D}_\varepsilon(q) = \{p : |q - p| < \varepsilon\}$.*

Proof. See appendix. □

[4]The following examples are due to Staël von Holstein (1970, p. 69).

The proposition says that every strictly proper scoring rule is locally concave around $p = q$. Our proof, which is presented in the appendix, is based on the Schervish (1989) representation of scoring rules. Intuitively, the proposition is closely related to the second-order condition for a maximum of $S(q,p)$ at $p = q$.

Figure 1.1 illustrates Proposition 1.2.2 for the spherical score. The vertical axis depicts different true distributions defined by $q \in (0,1)$. The horizontal axis depicts different candidate predictions indexed by p. The white area of the graph corresponds to the area in which $\frac{\partial^2 S(q,p)}{\partial p^2} \leq 0$; S is concave in this area (c.f. Appendix B). As implied by Proposition 1.2.2, the white area contains the main diagonal which has $p = q$. However, the white area extends far beyond this diagonal. That is, the admissible deviation ε between p and q (c.f. Proposition 1.2.2) is fairly large in this example.[5]

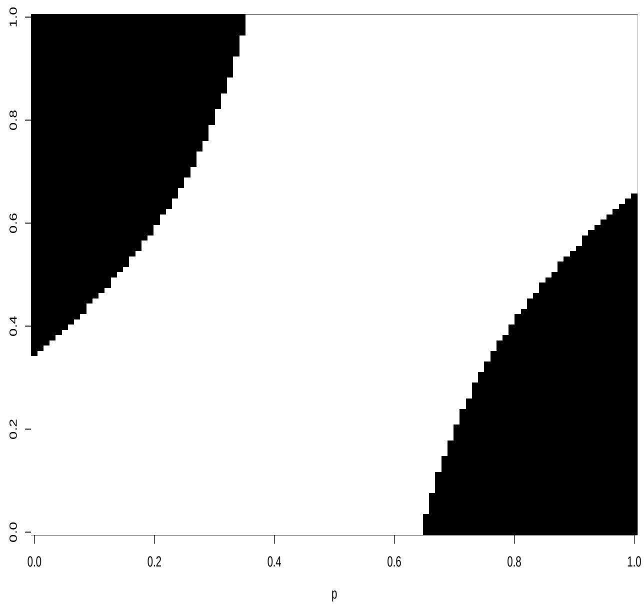


Figure 1.1: Illustration of concavity for the spherical score

The white area of the graph indicates that $\frac{\partial^2 S(q,p)}{\partial p^2} \leq 0$, for the spherical score and given values of q (vertical axis) and p (horizontal axis). The black area indicates that $\frac{\partial^2 S(q,p)}{\partial p^2} > 0$.

[5]Note that analogous graphs for the logarithmic and Brier scores would be completely white, since for these scoring rules it holds that $\frac{\partial^2 S(q,p)}{\partial p^2} \leq 0$ for all p and q between zero and one.

1.3 Implications

We next give a brief account of the relationship between concavity and the performance of linear combinations of probability forecasts. Originally proposed by Stone (1961), such combined forecasts have received considerable attention in recent years; see e.g. Wallis (2005), Ranjan and Gneiting (2010), Gneiting and Ranjan (2011), Clements and Harvey (2011) and Geweke and Amisano (2011).

Given a set of m probability forecasts $p_j, j = 1, \ldots, m$, a *linear pool* is given by

$$p_c = \sum_{j=1}^{m} \omega_j p_j,$$

with weights ω_j satisfying $0 \leq \omega_j \leq 1$ and $\sum_{j=1}^{m} \omega_j = 1$.

Proposition 1.3.1. *Consider m probabilistic forecasts $\{p_j\}_{j=1}^{m}$, with $p_j \in \mathcal{D} \subseteq [0, 1]$ f.a. j. Furthermore, suppose the scoring rule S is entirely concave on $\mathcal{D}$. Then,*

$$S(q, \sum_{j=1}^{m} \omega_j p_j) \geq \sum_{j=1}^{m} \omega_j S(q, p_j).$$

Proof. Follows from Jensen's inequality. □

Proposition 1.3.1 slightly generalizes the results by Staël von Holstein (1970, p. 69) who presents similar inequalities for specific scoring rules.[6] Clearly, the proposition does *not* exclude the possibility that one of the individual models attains a higher expected score than the combination; hence it does not say that combination is generally better than selection. Instead, the proposition is interesting because it defines a lower bound on the performance of the combination (relative to that of the individual models) which holds for any value of the unknown true probability q.

The value of the lower bound can be interpreted as the expected score of a forecaster who randomly selects one of the individual forecasts p_j, using the combination weights ω_j as selection probabilities.[7] While it is not common to explicitly randomize over forecasting models, selecting a single forecasting model

[6] Note that his results refer to realized, rather than expected, scores.
[7] An analogous interpretation also holds for the *realized* score (see Staël von Holstein, 1970, p. 69).

(e.g. via information criteria or based on historical forecasting performance) is common practice. The proposition suggests that the randomness inherent in such selection methods is harmful in terms of concave scoring rules.

We next show a partial converse to the proposition: If the expected score function is convex, forecast combination is generally dominated by the selection of a single model.

Proposition 1.3.2. *Suppose the scoring rule S and the true distribution q are such that S is convex on a set $\mathcal{D} \subseteq [0,1]$. Furthermore, consider m probabilistic forecasts $\{p_j\}_{j=1}^m$, with $p_j \in \mathcal{D}$ f.a. j, and $S(q, p_j) \neq S(q, p_k)$ f.a. $j \neq k$. Then, the optimal linear pool of these forecasts places full weight on a single model.*

Proof. From Jensen's inequality, we have that

$$\begin{aligned} S(q, \sum_{j=1}^m \omega_j p_j) &\leq \sum_{j=1}^m \omega_j S(q, p_j) \\ &< S(q, p_{j^*}), \end{aligned}$$

where $j^* = \arg \max_{j \in \{1,\ldots,m\}} S(q, p_j)$. Hence the (unique) optimum of the expected score function is attained by setting $\omega_{j^*} = 1$ and $\omega_j = 0$, f.a. $j \neq j^*$. □

Note that the assumptions of the proposition do not rule out the possibility that S is strictly proper. By Proposition 1.2.1, no strictly proper rule S can be entirely convex on $\mathcal{D}$. However, there may be some q such that S is convex on $\mathcal{D}$; c.f. the example of the spherical score in the last section.

Taken together, the two results in this section demonstrate the link between concave scoring rules and the performance of forecast combinations. Under concavity, the expected score of the combination satisfies a lower bound; Kascha and Ravazzolo (2010) note that this provides insurance against selecting a poor individual model. Under convexity, combination is generally inferior to selection.

1.4 Deciding when (not) to combine

1.4.1 Bounding the true probability

As shown in the last section, concave (convex) scoring rules set incentives to combine (select) forecasts. Some rules, like the logarithmic or Brier score, are entirely concave. For these rules, combined predictions will generally perform well relative to the individual models, in the sense of satisfying a lower bound. Other rules, like the spherical score, do have non-concave areas. That is, given a set $\mathcal{D}$ of forecasts, concavity of the scoring rule holds only for a certain range of true probabilities q. Knowing this range may be useful in practice: As a simple heuristic, a forecast user could decide to combine forecasts if he or she is willing to assume that q lies within this range. In the following, we illustrate the size of the range for an empirical application to predicting US recessions.

We consider a forecast user who faces two alternative predictions, p_1 and p_2. Given these predictions, we seek to identify the set of true distributions q which correspond to the concave area of the scoring rule. Formally, this set is defined as follows:

$$\mathcal{Q}(p_1, p_2) = \{q : S \text{ is concave on } \{p_1, p_2\}\}. \tag{1.3}$$

By the definition of concavity, $q \in \mathcal{Q}(p_1, p_2)$ implies that the combination satisfies a lower bound for *any* combination weight $\omega \in (0, 1)$ which is placed on the first model. This may be too strict a requirement. For example, the forecast user may only be interested in whether an equally weighted combination is promising. We therefore consider a second set which contains the true distributions under which the expected score of a combination exceeds the average of the expected scores *for a given combination weight* ω:

$$\tilde{\mathcal{Q}}(\omega, p_1, p_2) = \{q : \; S\left(q, \omega p_1 + (1-\omega)\; p_2\right) \geq \omega S(q, p_1) + (1-\omega) S(q, p_2)\}. \tag{1.4}$$

The following proposition summarizes some features of the two sets.

Proposition 1.4.1. *Assume that the scoring rule S is strictly proper. Then, the two sets defined in (1.3) and (1.4) have the following properties.*

1. $\mathcal{Q}(p_1, p_2) = \bigcap\limits_{\omega \in (0,1)} \tilde{\mathcal{Q}}(\omega, p_1, p_2)$
2. *The set $\mathcal{Q}(p_1, p_2)$ is an interval.*

3. *For given* $\omega \in (0,1)$*, the set* $\tilde{\mathcal{Q}}(\omega, p_1, p_2)$ *is a non-empty interval.*

4. *If* S *is entirely concave on* $[0,1]$*, then* $\mathcal{Q}(p_1, p_2) = \tilde{\mathcal{Q}}(\omega, p_1, p_2) = [0,1]$ *for any* $\omega \in (0,1)$.

Proof. See appendix. □

1.4.2 Empirical example

Below we illustrate the sets for a classical probabilistic forecasting problem in economics. Following Rudebusch and Williams (2009), we consider two alternative forecasts of the probability of a decline in quarterly US real GDP. Historically, this event has been a close proxy for a "recession", as defined by the National Bureau of Economic Research (NBER). The first forecast is the (average) expert forecast from the Survey of Professional Forecasters (SPF) run by the Federal Reserve Bank of Philadelphia. The second forecast is from a simple Probit model based on the term spread, which is the yield to a ten-year US government bond minus the yield to a three-month bill. Our empirical analysis follows Rudebusch and Williams (2009) in all implementation aspects. However, we extend their evaluation sample through the third quarter of 2011. Furthermore, for brevity we focus on forecasts made for the current quarter ("nowcasts"). In terms of forecast evaluation, we focus on the spherical score as the leading example of a strictly proper scoring rule which has non-concave areas. A detailed description of how we compute the forecasts, as well as the sets $\mathcal{Q}(p_1, p_2)$ and $\tilde{\mathcal{Q}}(\omega, p_1, p_2)$, is contained in the appendix.

The blue shaded area in Figure 1.2 displays the set $\mathcal{Q}(p_{1t}, p_{2t})$, across different quarterly evaluation periods t ranging from 1969Q1 to 2011Q3. As implied by the second result of Proposition 1.4.1, the set is connected at every period t, without any "white holes" in between. Over most of the evaluation period, the set extends well beyond the range of the two forecasting models. Hence the spherical score is concave for a broad range of possible true probabilities. The set $\tilde{\mathcal{Q}}(\omega, p_{1t}, p_{2t})$, for $\omega = 0.5$, can also be read off Figure 1.2. It consists of the shaded blue area, plus the adjacent green area. The size of the green area is rather modest in this example, suggesting that the relatively good performance of combinations is not restricted to a particular value of the combination weight.

Figure 1.2 suggests that during the entire evaluation period, forecast combination is a promising strategy for a large range of true probabilities. Thereby, the adjective "promising" is from an ex ante perspective, as it relates to the *expected* score of the combination. A natural next step is to evaluate the ex post performance of the combination, in terms of *realized* scores over the evaluation sample. Table 1.1 provides such an evaluation, using the equally weighted combination for simplicity. In addition to average spherical scores over the evaluation sample, the table reports Diebold and Mariano (1995) tests for equal predictive ability in terms of the spherical score. The table shows two main results: First, the SPF forecast significantly outperforms the Probit forecast but not the combination, using a 5% significance level and a two-sided test. Second, the combination significantly outperforms the Probit forecast.

We view these results as very positive for the equally weighted combination. Although the SPF turns out to be a better individual forecast than the Probit, it is questionable whether a forecast user would have selected the SPF forecast in real time. After all, the good performance of the SPF is known only after the end of the evaluation period.[8] In contrast, the good performance of the combination is in line with the theoretical analysis presented above, and could have been anticipated in real time. It is hence remarkable that the combination is statistically indistinguishable from the ex post better model (SPF), while being superior to the ex post worse model (Probit).

[8]Of course, a forecast user might have recognized the better performance of the SPF after (say) the first $\tilde{T}$ periods, and might have selected the SPF forecast on the basis of this information. However, this selection strategy has two drawbacks: First, it is sensitive to the choice of $\tilde{T}$, and we are not aware of a formal way of choosing this tuning parameter. Second, the strategy does not prescribe a forecast for the first $\tilde{T}-1$ evaluation periods.

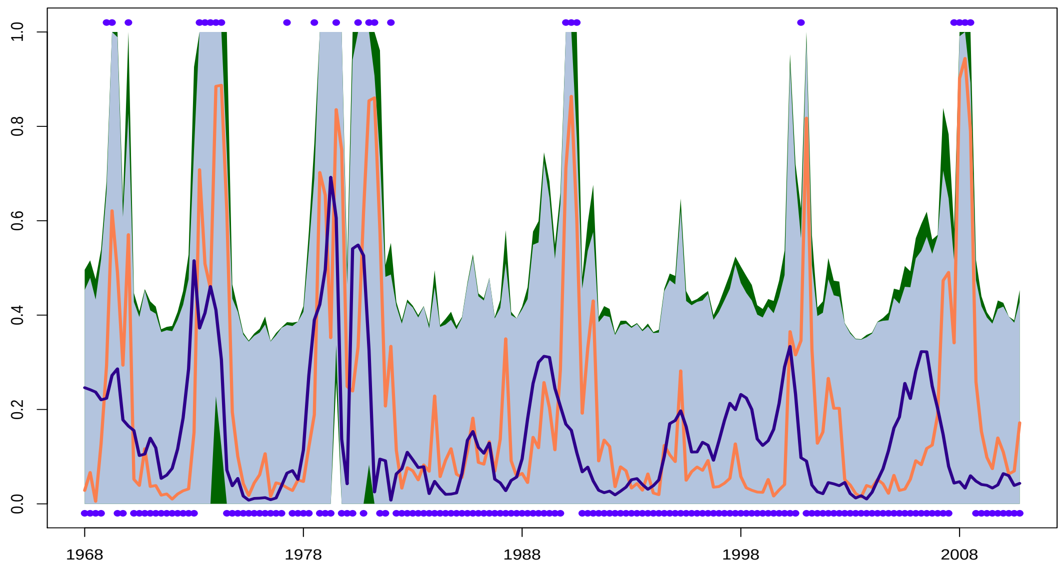


Figure 1.2: Concave areas in the empirical example

The orange and blue lines are SPF and Probit current-quarter probability forecasts of a decline in US real GDP, for an evaluation sample ranging from 1969Q1 to 2011Q3. Dots on the top of the graph indicate that a decline actually occurred; dots at the bottom indicate that there was no decline. The blue shaded area represents the set $\mathcal{Q}(p_{1t}, p_{2t})$ defined in Equation (1.3). The adjacent green area represents the set $\tilde{\mathcal{Q}}(\omega, p_{1t}, p_{2t}) \backslash \mathcal{Q}(p_{1t}, p_{2t})$, where $\omega = 0.5$ and $\tilde{\mathcal{Q}}(\omega, p_{1t}, p_{2t})$ is defined in Equation (1.4). Both sets refer to the spherical score.

	SPF	Probit	Combination
Avg. Spherical Score	0.926	0.882	0.918
EPA test vs Probit	$2.006^{0.045}$	•	•
EPA test vs Combination	$0.839^{0.402}$	$-2.700^{0.007}$	•

Table 1.1: Empirical forecast evaluation

The first row reports average spherical scores over the evaluation sample (1968Q4 to 2011Q3). The second and third rows report Diebold and Mariano (1995) test statistics for equal predictive ability (EPA). Superscript numbers are asymptotic standard normal p-values (two-sided test). A lag length of four is used in the Newey and West (1987) variance estimator which enters the test statistic.

1.5 Discussion

Focusing on a binary prediction problem, this paper has shown that all smooth strictly proper scoring rules are locally concave around the true event probability. This finding generalizes earlier research which has focussed on specific scoring rules. It implies that simple linear forecast combinations can be expected to perform well under a wide range of scoring rules, provided that the individual models are sufficiently close to the true probability.

These findings imply that the good performance of linear forecast combinations is partly an inevitable consequence of using a “reasonable” scoring rule. This interpretation contrasts the one of the extant literature which tends to view the good performance of forecast combinations as a purely empirical result; see also the comments by McNees (1992).

Finally, our statement that “combinations perform well relative to their components” is motivated by the existence of a lower bound on the expected score of the combination. Clearly, this type of analysis is not the only way to look at the problem. In particular, the lower bounds are not informative about the *efficiency* of combinations, i.e. their performance relative to the true model. Similar to Elliott (2011) who considers combinations of point forecasts, such analysis could deliver further insight into the performance of combined probability forecasts.

Appendix 1.A Proofs

Proof of Proposition 1.2.1

Proof. Assume that S is both strictly proper and entirely convex on $\mathcal{D}$, and let the true distribution be given by $q = \lambda^* p_1 + (1-\lambda^*)p_2$, with $p_1 \neq p_2 \in \mathcal{D}, \lambda^* \in (0,1)$. Then it holds that

$$\begin{aligned} S(q,q) &= S\left(q, \lambda^* p_1 + (1-\lambda^*)p_2\right) \\ &\overset{ent.convex}{\leq} \lambda^* S(q,p_1) + (1-\lambda^*)S(q,p_2) \\ &\overset{str.proper}{<} \lambda^* S(q,q) + (1-\lambda^*)S(q,q) \\ &= S(q,q). \end{aligned}$$

From the contradiction $S(q,q) < S(q,q)$ it follows that S cannot be both strictly proper and entirely convex on $\mathcal{D}$. □

Proof of Proposition 1.2.2

Proof. By the Schervish (1989) representation (c.f. Gneiting and Raftery, 2007, Theorem 3), a strictly proper scoring rule S satisfies

$$\begin{aligned} g_1(p) &= g_1(1) - \int_p^1 (1-c)\nu(c)dc \\ g_0(p) &= g_0(0) - \int_0^p c\nu(c)dc, \end{aligned}$$

where $\nu(c)$ is strictly positive on the unit interval and, by Assumption 1.2.1, is continuously differentiable. Leibniz' formula yields

$$\begin{aligned} g_1'(p) &= (1-p)\nu(p) \\ g_0'(p) &= -p\nu(p), \end{aligned}$$

so that

$$\begin{aligned} \frac{\partial S(q,p)}{\partial p} &= q g_1'(p) + (1-q) g_0'(p) \\ &= (q-p)\nu(p). \end{aligned}$$

Differentiating with respect to p gives

$$\frac{\partial^2 S(q,p)}{\partial p^2} = -\nu(p) + (q-p)\nu'(p). \tag{1.5}$$

Hence,

$$\left.\frac{\partial^2 S(q,p)}{\partial p^2}\right|_{p=q} = -\nu(q) < 0. \tag{1.6}$$

Since $S(q,p)$ is assumed to be twice continuously differentiable, Equations (1.6) and (1.5) imply that $S(q,p)$ is locally concave around $p=q$. □

Proof of Proposition 1.4.1

Proof. Ad 1): Follows directly from the definition of concavity.

Ad 3): By strict propriety of S, $q^* = \omega p_1 + (1-\omega)p_2$ is always contained in $\tilde{\mathcal{Q}}(\omega, p_1, p_2)$, so that the set is non-empty. Since $\tilde{\mathcal{Q}}(\omega, p_1, p_2)$ is a subset of $[0,1]$, the statement that the set is an interval is equivalent to the set being convex. We show convexity in the following. If q^* is the only element of $\tilde{\mathcal{Q}}(\omega, p_1, p_2)$, then the set is trivially convex. If there are at least two elements, denote two distinct elements by q^1, q^2. Consider $q^c = \mu q^1 + (1-\mu)q^2, \mu \in (0,1)$. For any p, it holds that $S(q^c, p) = \mu S(q^1, p) + (1-\mu)S(q^2, p)$. Hence,

$$\begin{aligned}
& S(q^c, \omega p_1 + (1-\omega)p_2) \\
= \ & \mu S(q^1, \omega p_1 + (1-\omega)p_2) + (1-\mu)S(q^2, \omega p_1 + (1-\omega)p_2) \\
\geq \ & \mu\{\omega S(q^1, p_1) + (1-\omega)S(q^1, p_2)\} + (1-\mu)\{\omega S(q^2, p_1) + (1-\omega)S(q^2, p_2)\} \\
= \ & \omega S(q^c, p_1) + (1-\omega)S(q^c, p_2),
\end{aligned}$$

which implies that $q^c \in \tilde{\mathcal{Q}}(\omega, p_1, p_2)$, and hence the set is convex.

Ad 2): Follows from i), iii), and the fact that the intersection of convex sets is again convex.

Ad 4): Follows directly from Definition 3. □

Appendix 1.B Details on Section 1.4

Data and forecasts

As mentioned earlier, we follow Rudebusch and Williams (2009) in all implementation aspects. We refer to their paper, as well as the online appendix available at

http://amstat.tandfonline.com/toc/jbes/27/4

for details.

Quarterly forecasts from the SPF is available on the Philadelphia Fed's website at

www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/anxious-index/

Real-time data on output growth is also available from this website:

www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/ROUTPUT/

We use the quarterly vintages in our analysis. Data on the term spread is taken from the FRED database run by the St. Louis Fed:

http://research.stlouisfed.org/fred2/

The term spread is constructed as the yield to a ten-year US government bond (FRED series GS10), minus the yield to a three-month bond (FRED series TB3MS). We construct quarterly data by averaging over three monthly rates.

A "decline in real GDP in quarter t" is defined as a decline in the level of real GDP in quarter t, as compared to quarter $t-1$. Following Rudebusch and Williams (2009), we use "first-final" data vintages to measure real GDP. In the following, we use the binary variable R_t, with $R_t = 1$ ($R_t = 0$) indicating that there was a (no) decline in quarter t.

The SPF forecast is computed as the average probability forecast over all participants of the survey. The current-quarter forecast ("nowcast") for quarter t is made in the middle of the quarter; this corresponds to $h = 0$ in the notation of Rudebusch and Williams (2009). The SPF forecasts can be used directly, without further statistical processing.

The Probit forecasts stem from a regression of R_t on a constant and the term spread S_{t-1}, i.e. the term spread lagged by one quarter. This lag is due to the fact that data from the current quarter is not yet available in the middle of the quarter, when the SPF forecasts are submitted. The Probit model is recursively re-estimated each period, using the latest available real-time data. Hence the Probit forecast is formally given by

$$\hat{P}(R_t = 1) = \Phi(\hat{\alpha}_t + \hat{\beta}_t S_{t-1}),$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. $\hat{\alpha}_t$ and $\hat{\beta}_t$ are estimated based on data available in the middle of quarter t, using maximum likelihood. All computations were done using the R programming language (R Development Core Team, 2012).

Determination of the sets

Figure 1.2 in Section 1.4 presents the sets $\mathcal{Q}(p_{1t}, p_{2t})$ and $\tilde{\mathcal{Q}}(\omega, p_{1t}, p_{2t})$, given the two forecasts p_{1t} and p_{2t} at date t. We next describe how to determine the two sets.

The following result establishes a simple method to obtain a conservative estimate of the first set, $\mathcal{Q}(p_{1t}, p_{2t})$, under the spherical score. For simplicity, we use the short notation $S_{pp}(q, \tilde{p}) := \frac{\partial^2 S(q,p)}{\partial p^2}\Big|_{p=\tilde{p}}$ in the following.

Proposition 1.B.1. *Consider the set $\mathcal{D}_t = \{p_{1t}, p_{2t}\}$; without loss of generality, let $p_{1t} \leq p_{2t}$. Then, for a given $q \in [0, 1]$, the spherical score is concave on $\mathcal{D}_t$ if $S_{pp}(q, p_{1t}) \leq 0$ and $S_{pp}(q, p_{2t}) \leq 0$.*

Proof. For general p, we have that

$$S_{pp}(q, p) = \frac{4p^2 - p(6q + 1) + 3q - 1}{((1 - p)^2 + p^2)^{2.5}},$$

which is negative iff the numerator is negative. This numerator is a strictly convex function of p. Hence for any $\mu \in (0, 1)$,

$$S_{pp}(q, \mu p_{1t} + (1 - \mu)p_{2t}) < \mu \underbrace{S_{pp}(q, p_{1t})}_{\leq 0} + (1 - \mu) \underbrace{S_{pp}(q, p_{2t})}_{\leq 0} \leq 0.$$

This implies that $S_{pp}(q, \tilde{p}) \leq 0$ for all $\tilde{p} \in [p_{1t}, p_{2t}]$. Hence, the spherical score is concave on $[p_{1t}, p_{2t}] \supseteq \mathcal{D}_t$, and thus also on $\mathcal{D}_t$. □

That is, a negative second derivative of $S(q, p)$ at $p = p_{1t}$ and $p = p_{2t}$ is a sufficient condition for concavity on $\{p_{1t}, p_{2t}\}$. Note that this result is visualized by Figure 1.1: For a given true probability q (vertical axis), the white area on the horizontal axis of the graph is an interval (rather than being disrupted by "holes").

We use the result in Proposition 1.B.1 to obtain a first estimate of the set $\mathcal{Q}(p_{1t}, p_{2t})$. This estimate (say, $\hat{\mathcal{Q}}(p_{1t}, p_{2t})$) is given by the values of q for which the sufficient condition of the proposition is satisfied. This estimate is conservative (i.e., $\hat{\mathcal{Q}}(p_{1t}, p_{2t}) \subseteq \mathcal{Q}(p_{1t}, p_{2t})$), since the condition of the proposition is not necessary. To determine $\mathcal{Q}(p_{1t}, p_{2t})$ more precisely, we then check concavity for candidate points $q \in \left([0, 1] \setminus \hat{\mathcal{Q}}(p_{1t}, p_{2t})\right)$ numerically. For given q, this is done by going through a fine grid of values $\lambda = 0.00001, ..., 0.99999$. If the inequality of Definition (2) is satisfied for each value of λ, we assert that $q \in \mathcal{Q}(p_{1t}, p_{2t})$; if not, we assert that q is not in the set.

The set $\tilde{\mathcal{Q}}(\omega, p_{1t}, p_{2t})$ is determined numerically. This is computationally much simpler than for the first set: For a given q, only a single value of $\lambda = \omega = 0.5$ has to be checked.

Bibliography

BRIER, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78, 1–3.

CLEMENTS, M. P. AND D. I. HARVEY (2011): "Combining Probability Forecasts," *International Journal of Forecasting*, 27, 208–223.

DIEBOLD, F. X., T. A. GUNTHER, AND A. TAY (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

DIEBOLD, F. X. AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

ELLIOTT, G. (2011): "Averaging and the Optimal Combination of Forecasts," *Working Paper*.

GEWEKE, J. AND G. AMISANO (2011): "Optimal prediction pools," *Journal of Econometrics*, 164, 130–141.

GNEITING, T., F. BALABDAOUI, AND A. RAFTERY (2007): "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society*, 69, 243–268.

GNEITING, T. AND E. A. RAFTERY (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

GNEITING, T. AND R. RANJAN (2011): "Combining Predictive Distributions," *Working Paper*.

GOOD, I. J. (1952): "Rational Decisions," *Journal of the Royal Statistical Society*, 14, 107–114.

HENDRICKSON, A. D. AND R. J. BUEHLER (1971): "Proper Scores for Probability Forecasters," *The Annals of Mathematical Statistics*, 42, 1916–1921.

KASCHA, C. AND F. RAVAZZOLO (2010): "Combining inflation density forecasts." *Journal of Forecasting*, 29, 231–250.

LARRICK, R. P. AND J. B. SOLL (2006): "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle," *Management Science*, 52, 111–127.

Lichtendahl, K. C., Y. Grushka-Cockayne, and R. L. Winkler (2012): "Is it Better to Average Probabilities or Quantiles?" *Management Science (forthcoming).*

Manski (2010): "When Consensus Choice Dominates Individualism: Jensen's Inequality and Collective Decisions Under Uncertainty," *Quantitative Economics*, 1, 187–202.

McNees, S. K. (1992): "The Uses and Abuses of "Consensus" Forecasts," *Journal of Forecasting*, 11, 703–710.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

R Development Core Team (2012): "R: A Language and Environment for Statistical Computing," *http://www.R-project.org.*

Ranjan, R. and T. Gneiting (2010): "Combining Probability Forecasts," *Journal of the Royal Statistical Society*, 72, 71–91.

Rudebusch, G. D. and J. C. Williams (2009): "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve," *Journal of Business and Economic Statistics*, 27 (4), 492–503.

Schervish, M. J. (1989): "A General Method for Comparing Probability Assessors," *Annals of Statistics*, 4, 1856–1879.

Shuford, E., A. Albert, and H. E. Massengill (1966): "Admissible probability measurement procedures," *Psychometrika*, 31, 125–145.

Staël von Holstein, C. A. S. (1970): *Assessment and Evaluation of Subjective Probability Distributions*, Economic Research Institute, Stockholm School of Economics; Norstedt (distr.).

Stone, M. (1961): "The Opinion Pool," *The Annals of Mathematical Statistics*, 32, 1339–1342.

Toda, M. (1963): "Measurement of Subjective Probability Distributions," *Division of Mathematical Psychology, Institute for Research, State College, Pennsylvania*, Report No. 3.

Wallis, K. (2005): "Combining Density and Interval Forecasts: A Modest Proposal," *Oxford Bulletin of Economics and Statistics*, 67, 983–994.

Chapter 2

Averaging across Asset Allocation Models

2.1 Introduction

It is found that (equally weighted) combination of models performs surprisingly well. This empirical finding holds in many areas including forecasts (see Smith and Wallis, 2009 and references therein), experts recommendations (Genre et al., 2013), assets (DeMiguel et al., 2009b) and others (see Clemen, 1989 for an excellent review). The average opinion appears to be constantly among the best. Surowiecki (2004) calls it the "Wisdom of Crowds". Mainly three explanations have been suggested. Bates and Granger (1969) argue that different models are based on different information sets. The combination of models helps to combine those information sets. Diebold and Pauly (1987) argue that models are differently affected by structural breaks. For example, models based on a longer past time period are stable but need longer to adapt to structural breaks than models with shorter memories. Stock and Watson (2004) argue that the true data generating process is more complex and of a higher dimension than even the most flexible models. The combined model is robust to misspecification of individual models.

Generally, the term diversification represents the investment in various assets across different asset classes, industries or regions. The ideosyncratic risk of individual assets is reduced by a diversification in assets. This paper is not about diversification in assets but about diversification in models. Each model has some estimation risk. The estimation risk can be reduced by model diversification (or model combination). Note that, both types of diversification are not connected. Model combination can lead to a more but also to a less diversified portfolio.

So far, the combination of multiple asset allocation models has not been analyzed. We conduct an empirical study using five different asset allocation models of three different model classes. We apply six different data sets. As common performance functions we use the Certainty Equivalent, the Sharpe Ratio, the Variance, the Value at Risk and the Expected Shortfall. We find that no model constantly outperforms the others. The best model depends on the risk function and the data set. Even for a given risk function and a given data set, the ranking of models changes over time. The presented argumentation of the forecasting literature seems to hold for asset allocation models as well. The models are subject to different information sets, structural breaks and a model capturing the true data generating process does not exist.

In a seminal paper Markowitz (1952) introduced a fundamental concept of portfolio optimization. The approach is convincing in theory, but when it comes to empirical applications it turns out that the estimation of portfolio weights is challenging (Britten-Jones, 1999). Most models are based on the first two moments of the return process. Frost and Savarino (1986) showed that in particular the mean is difficult to estimate. Also the estimation error in the covariance matrix is large (Chan et al., 1999). In times when the estimation error is large, naive models such as equally weighting and value weighting perform well. In stable periods, estimation is valuable and models conditioned on the moments of the return perform well. In light of the model uncertainty Tu and Zhou (2011) combine the tangency strategy with the equally weighted portfolio. The idea is also common to shrinkage approaches, which shrink one model with more estimation risk towards some stable target model (see e.g., Jorion, 1986). These shrinkage methods can be regarded as the combination of only two models.

Our approach goes further and combines multiple models and gives a theoretical support for combination of models.

Finding that the ranking of models changes over time, the combination of models stabilizes the performance. One might neither expect a remarkably good nor an exceptionally bad performance. The question arises why the combined model should perform better than individual models. We follow the argumentation of Lichtendahl et al. (2012) and Krueger and Schanbacher (2012) stating that under a concave performance function, the performance of the *combined model* is necessarily better than the *combined performance* of the individual models. Most common performance functions are concave. A large class of concave performance functions is the class of coherent performance (or risk) functions of Artzner et al. (1999). We find that three out of the five considered risk functions are always concave while the remaining two are concave under certain conditions. For concave risk functions and a situation of changing model ranking the combined model performs well *by definition*.

The paper contributes to the literature in two ways. First, we theoretically show that the combined model has to perform well with respect to the standard performance functions used in literature as well as by practitioners. Common performance functions are concave - at least in most practical applications. Second, we present a comprehensive empirical analysis using five standard models of three distinct model classes. The evaluation is performed on six different data sets with respect to five different performance functions. We find that there is no constantly dominating model. The performance depends on the risk function, the data set and the time considered. Under these circumstances ex-ante selection of the best individual model is difficult

while combining models is rewarding. The combined model can even outperform all individual models.
The remainder of the paper is structured as follows. Section 2.2 introduces the concept of concavity. Loss functions are presented and analyzed as well as conditions for concavity of all performance functions are stated. Section 2.3 presents the empirical study. We use common models, apply standard data sets and discuss the empirical results. Section 3.6 summarizes and concludes.

2.2 Portfolio Evaluation

Consider a market with $m \in \mathbb{N}$ asset allocation models. The (excess) returns of the models are given by the $m-$dimensional random vector $\mathbf{R} = (R_1, \ldots, R_m)$. In our analysis we restrict ourselves to the combination of asset allocation models. We consider linear combinations only. The weight vector is given by $\Lambda = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$. We assume that the investor is fully invested, i.e. $\iota'\lambda = 1$. Asset allocation models are designed for accumulating wealth rather than destroy wealth. If the investor believes a certain asset allocation approach destroys wealth he might include the inverse approach as an additional (wealth creating) model. Without restriction we suppose the investor is not short in an asset allocation model, e.g. $\lambda_i \geq 0$. In the following when speaking about weights λ, we imply that $\iota'\lambda = 1$ and $\lambda_i \geq 0$, unless noted otherwise. The return of the combined model is given by $\Lambda'\mathbf{R}$. A major aim in finance is to evaluate the performance of model. The following definition establishes some general objective function for the investor.

Definition 4. *Consider two returns $R, \tilde{R}$ with some pdf f and $\tilde{f}$, resp. Let the investor evaluate both returns based on some parameter(s) of interest $\theta = H(f)$ where $\theta \in \Theta$ and $H : \; f \mapsto \theta$ (similar for $\tilde{\theta} = H(\tilde{f})$). Let there be a function $F : \Theta \to \mathbb{R}$. The investor prefers return R over return $\tilde{R}$ if and only if $F(\theta) > F(\tilde{\theta})$. Then function F is called a performance function.*

For simplicity we write $F(R)$ instead of $F(H(f))$ with f being the pdf of R. As F should be maximized it can be regarded as the "utility" of the investor. An important special case is an investor evaluating returns based on the first two moments, i.e. $\theta = (\mu, \sigma^2)$ with $\mu = \int_{\mathbb{R}} x f(x) dx$ and $\sigma^2 = \int_{\mathbb{R}} (x-\mu)^2 f(x) dx$. In this case, his performance function might be the Sharpe ratio, i.e. $F(\mu, \sigma^2) = \mu / \sqrt{\sigma^2}$. An important property of a performance function is concavity introduced by the following definition.

Definition 5. *A function $F : \mathcal{A} \to \mathbb{R}$ is called concave if*

$$\lambda F(x) + (1-\lambda)F(\tilde{x}) \leq F\left(\lambda x + (1-\lambda)\tilde{x}\right) \tag{2.1}$$

for any $x \neq \tilde{x} \in \mathcal{A}$ and $\lambda \in (0,1)$. F is called strictly concave if equality 2.1 holds strictly.

Artzner et al. (1999) argue that a coherent performance function should be superadditive[1] (i.e. $F(R+\tilde{R}) \geq F(R) + F(\tilde{R})$) and positive homogenous (i.e. $F(\lambda R) = \lambda F(R)$ f.a. $\lambda \geq 0$). Each coherent performance function is therefore concave. As the risk of a position might increase nonlinearly the opposite is not necessary true. The motivation of coherent performance functions is that diversification should not increase the risk. The definition of concavity is based on the combination of two elements only. The following lemma helps to generalize the effects of a concave function.

Lemma 2.2.1. *For concave $F : \mathcal{A} \to \mathbb{R}$ it holds that*

$$\sum_{i=1}^{m} \lambda_i F(x_i) \leq F\left(\sum_{i=1}^{m} \lambda_i x_i\right)$$

with weights $\Lambda = (\lambda_1, \ldots, \lambda_m)$ and $x_i \in \mathcal{A}$.

Proof. see 2.A.1 □

Most applied work takes the objective function as given. Either because of calculus reasons (e.g. MSE) or traditional reasons (e.g. Sharpe Ratio). The properties of the performance function are rarely discussed. The main driver of diversification is the concavity of the applied performance function. The following Theorem 2.2.1 is a direct consequence of the definitions and the lemma above. Theorem 2.2.1 is simple but its implications are often neglected.

Theorem 2.2.1. *An investor applies some concave performance function F. Let m asset allocation models generate return $\mathbf{R} = (R_1, \ldots, R_m)$. Then the investor prefers a combination of the models over the combined performance of the models, i.e.*

$$\sum_{i=1}^{m} \lambda_i F(R_i) \leq F\left(\sum_{i=1}^{m} \lambda_i R_i\right) \tag{2.2}$$

[1] In their paper Artzner et al. (1999) require subadditivity as they minimize (not maximize) their loss function.

for all weights $\Lambda = (\lambda_1, \ldots, \lambda_m)$.

Proof. The Theorem follows from the concavity property and lemma 2.2.1. □

Despite its simplicity the conclusions drawn by Theorem 2.2.1 are rich. Inequality 2.2 does not hold in probability. Instead it shows that the investor combining asset allocation models *necessarily* outperforms the combined performance of the individual assets.

To emphasize the statement of Theorem 2.2.1 consider the following special case. Suppose there are several investors $i = 1, \ldots, m$ sharing utility function U. Each investor i applies his (subjectively best) asset allocation model generating return R_i. Holding the average portfolio of the investors leads to the average return $\overline{r} = \frac{1}{m}\sum_{i=1}^{m} r_i$. But the corresponding utility U will be *greater* than the average utility $\overline{U} = \frac{1}{m}\sum_{i=1}^{m} U(r_i)$ of all investors[2]. Investing in the average portfolio makes one better off than the average investor.

An alternative idea to look at Theorem 2.2.1 is the following. Facing m different models that generate returns $\mathbf{R}$ one could either combine models or select a favorite model. Theorem 2.2.1 says that in expectation one is better off by taking the equally weighted asset allocation model ($\Lambda = \frac{1}{m}\iota$) than selecting randomly one individual model. Applying a concave performance function leads to a small gain, if a "good" model is chosen but a large loss if a "bad" model is selected.

One might get the impression that the performance of the combined model lies in between the best and the worst individual performance. The combined model is never worse than the worst individual model. However, it can be better than the best individual model. A brief example shall highlight this fact. Let returns of two (not perfectly correlated) models r_1, r_2 be both $N(\mu, \sigma^2)$ distributed. Than the individual variance is given by $V(r_1) = V(r_2) = \sigma^2$. The variance of the averaged model is given by $V\left(\frac{r_1+r_2}{2}\right) = \frac{1}{4}\left(\sigma^2 + \sigma^2 + 2\rho\sigma^2\right) < \sigma^2$. The variance of the averaged model is therefore smaller than the smallest variance of all individual models. The mean for the individual models and the averaged model is the same, i.e. μ. In this setting, *any* sensible performance function should prefer the averaged model over both individual models. The example seems to be artificially constructed. But in reality similar situations are likely to occur. The empirical study in section 2.3 underlines that in most situations there is no model which constantly outperforms its competitors. Generally some models (e.g. conservative investment strategies) perform better in certain periods (e.g. crises) while performing bad compared to

[2]We assume U to be strictly concave and $r_i \neq r_j$ for some $i, j \in \{1, \ldots, m\}$.

their competitors in other periods. An average portfolio profits from gains but suffers disproportionately less from losses. It might even beat the ex-post best individual model. A general form of the stated example is given by the following corollary.

Corollary 2.2.1. *An investor applies some strictly concave performance function F. He faces m different asset allocation models. Model $i \in \{1, \ldots, m\}$ generates return R_i. Suppose the asset allocation models are of equal strength, i.e. $F(R_i) = F(R_j)$ for all $i, j \in \{1, \ldots, m\}$. Then any combined model is better than each individual model, i.e.*

$$F(R_j) < F\left(\sum_{i=1}^{m} \lambda_i R_i\right)$$

for all models j and weights $\Lambda = (\lambda_1, \ldots, \lambda_m)$.

Proof. Follows directly by Theorem 2.2.1. The inequality holds strictly as the performance function is strictly concave and the models are different. □

It is unrealistic that all models are of exact equal strength. But Corollary 2.2.1 improves our understanding of combined models. Under concave performance functions we have two results which at the first glance might be contradicting each other. Firstly, combined models are much better than the combined performance, if the performance of the models strongly differs. For large performance differences of the models, a combined model ensures that the investor does not pick an extremely "bad" model. Evaluating the investment strategies ex-post, a choice of an individual best model can be better. But ex-ante the future best model is not known. A strategy which ensures to be surely above average seems to be a magic strategy. If the model which will definitely perform best is known ex-ante, there is no need to apply models anymore. Reality can then be exactly described. Modestly stated, these situations are rare in social sciences. On the other hand, the ex-post best model is most likely better than the combined model. That leads us to the second consideration made by Corollary 2.2.1. When does a combined model even ex-post outperform the individual models? Most likely if the individual models show on average a similar performance but their performance changes over time. If an investor knows a set of good models but cannot determine ex-ante the best model, combination of models can outperform all individual models.

Having highlighted the power of Theorem 2.2.1, we turn to the assumptions to obtain the results. Note that inequality 2.2 is mechanic and technical. It holds always and is not stochastic in a sense that it holds in expectation or in probability only. Other

assumptions often made in literature are also neglected. Theorem 2.2.1 holds for all weights - not only for a certain combination of assets. No distributional assumptions about returns have to be made. The results advocating diversification also do not rely on some (negative) correlation structure of returns. For an investor who uses a concave performance function and does not know ex-ante a dominating model, averaging the models is in expectation better than randomly picking one.
Of all possible combinations of models, the averaged model combination plays the most important role. Assume a set of stable models is available. If the investor has no preference towards a specific model, he is advised to consider the average portfolio. With respect to a concave performance function in particular the equally weighted combination of all models will perform well. The requirement of concavity might be too strong. We introduce a weaker concept than concavity and consider the equally weighted combined model only.

Definition 6. *Let there be m asset allocation models which generate return $\mathbf{R} = (R_1, \dots, R_m)$ and some performance function F. Then we say F has dominance in average returns with respect to $\mathbf{R}$ if it holds that*

$$\frac{1}{m}\sum_{i=1}^{m} F\left(R_i\right) \leq F\left(\frac{1}{m}\sum_{i=1}^{m} R_i\right)$$

All concave performance functions have dominance in average returns but not vice versa. Theorem 2.2.1 and Corollary 2.2.1 can be stated similarly for performance functions with dominance in average returns. Under those performance functions the performance of the average portfolio is better than the average performance of the individual portfolios. Having defined the two concepts, concavity and dominance in average returns, we can start to analyze common performance functions.

2.2.1 Performance functions

Now we consider common performance functions used in literature as well as by practitioners. To highlight the effect of combined models, we introduce the performance functions using a weighted return, i.e. $w'\mathbf{R}$. Weight w can be either the weights of different asset allocation models Λ or the weight of different assets. The latter weights are usually denoted by w. To maintain the generality of the performance function we denote the weights by w. The first three performance functions which we introduce are the Certainty Equivalent, the Sharpe ratio and the Variance. These performance functions are based on the first two moments $\mu = E(\mathbf{R})$

and $\Sigma = Var(\mathbf{R})$. The final two performance functions are the Value at Risk and the expected Shortfall. Both are based on the tail distribution of $\mathbf{R}$.

Certainty Equivalent

The *Certainty Equivalent* is defined by

$$CE_\gamma(w'\mathbf{R}) = w'\mu - \frac{\gamma}{2}w'\Sigma w \tag{2.3}$$

Parameter γ reflects the investor's risk aversion with γ being large (small) for a risk-averse (risk-seeking) investor. For several reasons the Certainty Equivalent has become popular in literature (see e.g. Cho, 2011). The CE_γ covers a large range of potential investors. It includes the risk-neutral investor ($\gamma = 0$) and the global minimum variance investor ($\gamma = \infty$). It is also equivalent to the Markowitz (1952) portfolio optimization problem. Further it can be shown (see 2.B.1) that the investor maximizes the CE if
(i) his utility function is quadratic or
(ii) $\mathbf{R}$ is normal distributed and the investor's utility function is exponential or
(iii) the investment horizon is short.
The Certainty Equivalent is a concave performance function (see 2.A.2). If the returns have different variances the Certainty Equivalent is even strictly concave.

Sharpe Ratio

Originally introduced as the "reward-to-variability ratio" (Sharpe, 1964) the performance function is now commonly called the *Sharpe ratio* (1994) . It is defined by

$$SR(w'\mathbf{R}) = \frac{w'\mu}{\sqrt{w'\Sigma w}}$$

The Sharpe ratio measures return per unit risk (standard deviation). It is also known as the Sharpe Index (Radcliff, 1990) or the Sharpe Measure (Bodie et al., 1993). The actual definition of the Sharpe Ratio uses the excess return over a benchmark. Often the benchmark applied is the risk free rate. We are interested in comparing different models and set the benchmark to zero. Generally, the Sharpe ratio is no concave performance function and does not have dominance in average returns. Under certain conditions, however, the Sharpe ratio can be concave or at least obtain dominance in average returns. A sufficient condition for dominance in average returns is that risky models have a higher Sharpe ratio than less risky models. If $n = 2$, the condition even assures concavity (see 2.A.3 for a discussion and the proof).

Variance

As in particular the mean of returns is difficult to estimate (see Merton, 1980) one might consider the risk in form of the variance only. The performance function *MinVar* is defined by

$$MinVar(w'\mathbf{R}) = -w'\Sigma w$$

We consider the negative variance as our performance functions are positively oriented, i.e. the larger the better. In order to stick to common notation in our empirical study we present the variance instead of the negative variance. The MinVar can be regarded as a special case of the CE_γ with $\gamma \to \infty$. The MinVar is a concave performance function as the proof for the CE (2.A.2) still holds in the limit.

Value at Risk

The *Value at Risk* (VaR) is probably the most popular performance function in risk management. The Basel II Accord supported the VaR and defines it as the main measure of market risk. For a given significance level α, the VaR is defined as the minimum value that the loss exceeds with a probability not smaller than α.

$$VaR_\alpha(w'\mathbf{R}) = \inf\{x \in \mathbb{R} : Prob(w'\mathbf{R} \leq x) \geq \alpha\}$$

Essentially, the VaR is the $\alpha-$quantile of the return. Common significance values are $\alpha = 1\%$ or $\alpha = 5\%$. The time horizon is one month as it is the frequency of the returns. Artzner et al. (1997) show several counter-intuitive properties of the VaR. Generally the VaR is neither concave nor has dominance in average returns (see 2.A.4 for a counter-example). However, the VaR is a concave performance function if the return's VaR can be described by $VaR_\alpha(R_i) = \mu_i - \chi\sigma_i$ (see also 2.A.4 for a proof) with constant χ depending on the distribution, the degrees of freedom and confidence level α. This holds for example if returns are $t-$distributed.

Expected Shortfall

The *Expected Shortfall* (ES) is an alternative performance function to determine the downside risk of an asset. The ES is also known as the conditional value at risk, average value at risk or expected tail loss. It estimates the expected loss of a return below the VaR.

$$
\begin{aligned}
ES_\alpha(w'\mathbf{R}) &= \frac{1}{\alpha}\int_0^\alpha VaR_\gamma(w'\mathbf{R})d\gamma \\
&= E\left[w'\mathbf{R}|w'\mathbf{R} < VaR_\alpha(w'\mathbf{R})\right]
\end{aligned}
$$

Embrechts et al. (2002) show that the ES overcomes a number of difficulties which the VaR suffers from. In particular the ES is a concave performance function (see 2.A.5 or Foellmer and Schied, 2002).

2.2.2 Example

We learned that some performance functions are entirely concave, while others are not. Entirely concave measures do not necessarily prefer a combined model stronger than partly concave ones. Loss functions which are not concave might be (i) entirely concave for the given return structure $\mathbf{R}$ or (ii) still have strong dominance in average returns. We find that counter-examples for concavity of the VaR are mostly artificially constructed. For most return distributions the VaR is concave. The second point deals with the region of interest. No entire concavity is needed in order that a combined model performs well. The combined model should perform well for the applied weights. An investor averaging over all models is interested in dominance in average returns of the performance function; not in the concavity of the performance function. For a discussion on local concavity of performance functions we refer to Krueger and Schanbacher (2012).
To analyze the behavior of different performance functions suppose there are two models available. The returns of the risky model are $N(0.01, 0.005)$ distributed and the returns generated by the low-risk model are $N(0.005, 0.0025)$ distributed. Monthly returns of the models we consider follow roughly the stated distribution. Let the return of the models be correlated with $\rho = 0.7, 0, -0.3$. Figure 2.1 presents the distribution of returns of the individual models as well as the returns of their equally weighting combination. While the mean of the combined model stays the same (0.0075), negative correlation leads to a lower variance for the combined model. The highest diversification gain can be achieved for negative correlation between the models. Table 2.1 presents the effect of combination with respect to different performance functions. All performance functions are positively oriented with higher values being better. We find that a averaged model always outperforms the average

	CE	Sharpe	MinVar	VaR	ES
high-risk model	2.50	1.41	-5.00	-10.67	-3.94
low-risk model	1.25	1.00	-2.50	-7.75	-2.59
Aver	1.88	1.21	-3.75	-9.21	-3.27
Comb $\rho = 0.7$	2.06	1.25	-3.63	-9.18	-3.26
Comb $\rho = 0$	4.69	1.73	-1.88	-6.39	-2.56
Comb $\rho = -0.3$	5.81	2.24	-1.13	-4.78	-2.16

Table 2.1: Example: Averaged model
Equally weighted combination of $N(0.01, 0.005)$ and $N(0.005, 0.0025)$. CE and MinVar are scaled by 10^3, VaR and ES by 10^2 and Sharpe by 10.

performance of the individual models. That condition has to be satisfied for concave performance functions CE, Var and ES. For the given returns SR and VaR have dominance in average returns. We actually find that rewards of the combined model are highest for the non-concave performance functions as the Sharpe Ratio and the VaR. If the assets are uncorrelated or negative correlated (i.e. $\rho \leq 0$) the combined model even outperforms the best individual model. Note that the best model depends on the performance function. CE and Sharpe ratio select the high-risk model while the risk-averse measures MinVar, VaR and ES choose the low-risk model. An investor who is uncertain which performance function to choose, is even more advised to combine.
To closer analyze the effect of a flexible combination of models, Figure 2.2 shows the performance of a combination of models with weights $(\lambda, 1-\lambda)$, $\lambda \in (0,1)$. A twice continuous differentiable function is concave if and only if the second derivative is negative (Varian, 1992). We observe that all performance functions but the Sharpe Ratio are concave in this setting. For negative and zero correlation the Sharpe Ratio is concave in a large center part. In line with the argumentation of above an equally weighted combination of models is rewarding for all performance functions. For positive correlation $\rho = 0.7$ we find that all considered performance functions are entirely concave. In empirical work most models are correlated at about this level. Note that the degree of concavity is varying. The performance functions are stronger concave if the assets are negatively correlated. In economic terms, for negative correlated assets, one can hedge the risk better and therefore has a higher incentive to combine. Similar results can be obtained by combining t distributions. We conclude that in standard applications the common performance functions are concave or at least have dominance in average returns and reward model combination.

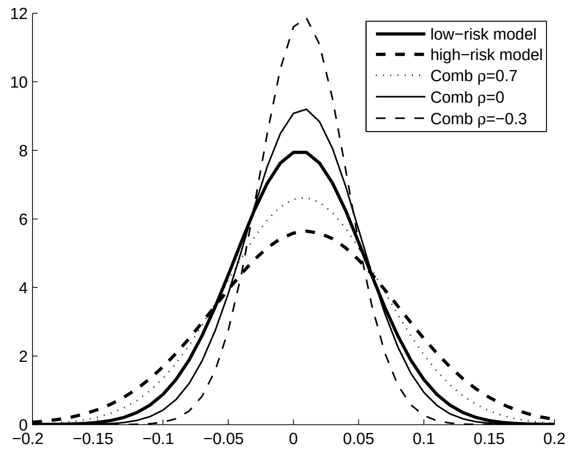


Figure 2.1: Example: Averaged model
Equally weighted combination of $N(0.01, 0.005)$ and $N(0.005, 0.0025)$ with correlation $\rho \in \{-0.3, 0, 0.7\}$.

2.3 Empirical Study of Model Combination

The result of Theorem 2.2.1 is only qualitative. The combined model performs at least as good as the combined performance of the individual weights. It is nothing claimed about the degree of the improvement. Inequality 2.2 depends strongly on the concavity of the performance function and returns $\mathbf{R}$. For highly concave performance functions and varying returns $\mathbf{R}$, the combined model strongly outperforms the combined individual investments. To analyze a quantitative improvement of model combination, one needs to run an empirical study for various performance functions and returns $\mathbf{R}$. The applied performance functions were introduced and analyzed in the section before. The models are introduced in section 2.3.1. The applied data sets are presented in section 4.4.2. The evaluation procedure is explained by section 2.3.3. Section 2.3.4 shows empirically that combination of asset allocation models is indeed rewarding.

2.3.1 Asset Allocation Models

In the following section we present the models for asset allocation which we use in our analysis. As the models are familiar to most readers, we only provide a brief overview. The reader who seeks further information or the derivation of the models is given the original references. We use models which we believe are among the most prominent ones in literature. We broadly distinguish between three different categories. To reduce estimation risk one can restrict models, use shrinkage approaches

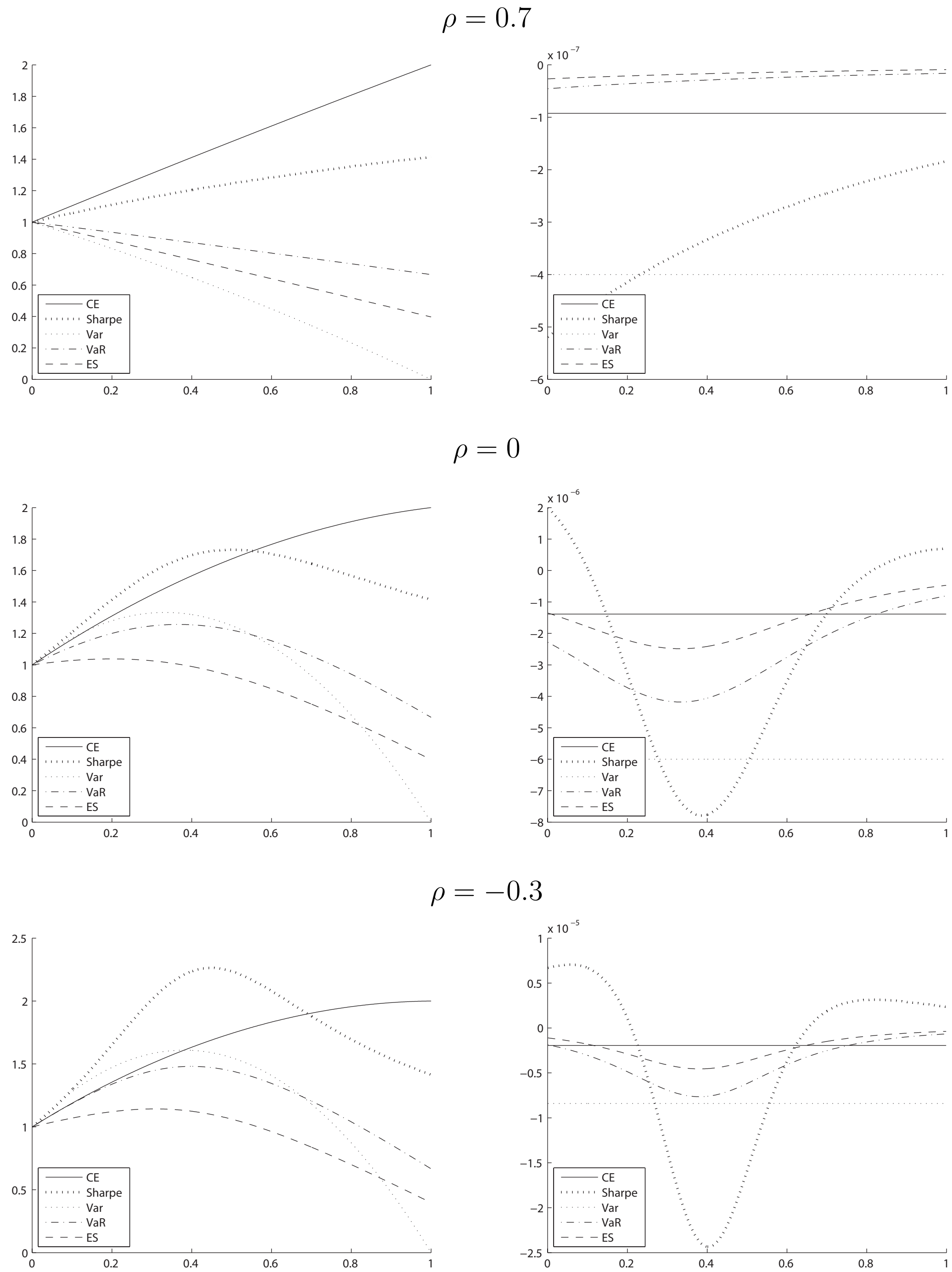


Figure 2.2: Concavity of performance functions.
Combination $\lambda R_1 + (1-\lambda)R_2$ with $\lambda \in [0,1]$ (x-axis). Here $R_1 \sim N(0.01, 0.005)$, $R_2 \sim N(0.005, 00.25)$ and correlation $\rho \in \{0.7, 0, -0.3\}$ (top down). Plots on the left side show (scaled) performance of the different performance functions. Plots on the right side show the second derivative (concavity).

or decide for naive strategies. Ignoring the mean and restricting short-selling are the most used constrained models. The most prominent shrinkage approach is the Ledoit and Wolf (2003, 2004a,b) portfolio. Naive approaches (equally weighted and value weighted) are independent of estimation risk. A summary of the used models is listed in Table 2.2. Most other models behave similarly to the considered ones or combinations of them.

Let $\mathbf{R}_t$ be an n dimensional return vector of n assets. In our empirical study t is ranging from 1 to T on a monthly frequence. The expected return vector and covariance matrix are given by μ_t and Σ_t. The sample counterparts are denoted by $\hat{\mu}_t$ and $\hat{\Sigma}_t$. The moments are estimated in a rolling window of length $h = 60$ months. The $n-$dimensional unity vector is given by ι while the $n \times n$ identity matrix is denoted by I.

In the following, we state the theoretical optimal weights for different approaches. The weights often rely on the first two moments of returns. In our empirical study we follow the classic "plug-in" approach: The individual approaches are solved by replacing the mean and variance by their sample counterparts $\hat{\mu}$ and $\hat{\Sigma}$, resp. For simplification we drop the time index t.

Minimum Variance

In the standard mean-variance framework, in particular the mean is difficult to estimate (see Merton, 1980 or Best and Grauer, 1991 for a sensitivity analysis). An alternative approach to minimize the portfolio variance only, i.e.

$$w_{MinVar} = \arg \min_{w, \iota w = 1} w' \Sigma w$$

The minimum variance portfolio (MinVar) can be regarded as the limiting case of equation 2.3 with $\gamma \to \infty$. The closed form solution is given by $w_{MinVar} = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}$. The investor either ignores the returns or restricts the expected returns to be equal, i.e. $\mu_i = \mu_j$ f.a. $i, j \in \{1, \ldots, n\}$. In practical application the portfolio can still behave instable if the estimated covariance matrix is almost singular.

MV with Short-selling Restriction

Large short positions are one of the main drivers for instable portfolio weights. Extreme positions can be circumvented by restricting short-selling. The short-selling restricted mean-variance portfolio (MVSR) is given by

$$w_{MVSR} = \arg \max_{w, \iota w=1, w_i \geq 0} w'\mu - \frac{\gamma}{2} w'\Sigma w$$

It can be shown that imposing the shortsale constraint is equivalent to either shrink the expected return towards the average or to shrink elements of the covariance matrix (Jagannathan and Ma, 2003). It can be also regarded as a special case of the approach of Fan et al. (2012) who constrain the portfolio weights. Their corresponding restriction binds the L_1 norm of the weights, i.e. $||w||_1 = 1$.

Ledoit-Wolf

So far, all proposed portfolio allocation models suffer of estimation risk of the covariance matrix. If the covariance matrix is almost singular, the weights are extreme and instable. Ledoit and Wolf (2003, 2004a,b) propose to estimate the covariance matrix by an optimally weighted average of two existing estimators (LW model). Namely the sample covariance matrix and a target correlation matrix F.

$$\Sigma_{LW} = \delta F + (1 - \delta)\Sigma$$

where shrinkage intensity δ is chosen as in Ledoit and Wolf (2004a). As the shrinkage target F one might use the single-factor model of Sharpe (1963) or a constant correlation covariance matrix. We apply the constant correlations approach and choose F by means of the sample variances and the average sample correlation, i.e. $f_{ii} = \Sigma_{ii}$, $f_{ij} = \frac{2\sqrt{\Sigma_{ii}\Sigma_{jj}}}{n(n-1)} \sum_{k=1}^{n-1} \sum_{k=k+1}^{n} \frac{\Sigma_{kl}}{\sqrt{\Sigma_{kk}\Sigma_{ll}}}$. The shrunken covariance matrix Σ_{LW} can be applied to obtain the minimum variance weights.

$$w_{LW} = \arg \min_{w, \iota' w=1} w'\Sigma_{LW} w$$

Alternative constraints such as bounds for the weights can be applied. However, there is no clear rule on how to select the bounds. Therefore additional constraints lead to arbitrary and incomparable results and are neglected in our analysis.

Equally Weighted

The equally weighted portfolio (EQ) is completely ignorant to the data and gives equal weight to all assets:

$$w_{Equal} = \frac{1}{n}\iota$$

The equally weighted portfolio can often be interpreted as a limit case of several optimization criteria. The equal weights can be obtained by optimizing the CE of eq. 2.3 under the constraint that $||w||_2 = \frac{1}{n}$ (see an asset allocation application of the ridge regression in DeMiguel et al., 2009a). It can be regarded as the Ledoit-Wolf weights with $F = I$ and $\delta = 1$. Or as an optimization with the constraint that the expected returns are proportional to total risk, $\mu \sim \Sigma\iota$.
Despite the total ignorance of data the equally weighted portfolio performs well compared to more sophisticated methods (DeMiguel et al., 2009b).

Value Weighted

In a CAPM framework the optimal strategy is given by a value weighted portfolio (Market). Each asset is given its market weight. We identify the benchmark market portfolio by investing at the beginning of the sample period the same amount in each asset. The portfolio is not restructured afterwards. Hence, the turnover of the value weighted strategy is zero. Value weighting is common for indicies such as the S&P500. Jorion (1991) finds that value-weighted portfolio have a similar performance to the equally weighted and the minimum-variance portfolio.

Combined Model

The final asset allocation model consists of a combination (Comb) of the previous introduced models. The weights of the combined model are given by

$$w_{Comb} = \frac{1}{m} \sum_{j=1}^{m} w^{(j)}$$

with $w^{(j)}$ denoting the weight of model $j \in \{MinVar, MVSR, LW, Equal, Market\}$ and $m = 5$ being the number of asset allocation models. In this paper we consider the equally weighted combination of asset allocation models only.

2.3.2 Data

To analyze the performance of the combined model, we consider a broad range of different data sets. A overview is given by Table 4.2. We show that combined model does not work well only in one specific field of asset allocation but leads to a general improvement independent of the particular asset classes. In all cases the data consists of monthly returns. The applied data sets are commonly chosen in literature (see

Model	Reference	Abbreviation
Restricted		
Minimum Variance	Merton (1980)	MinVar
MV without Short-selling	Jagannathan and Ma (2003)	MVSR
Shrinkage		
Ledoit-Wolf	Ledoit and Wolf (2004a)	LW
Naive		
Equally weighted	DeMiguel et al. (2009b)	Equal
Value weighted	common for indices (e.g. S&P 500)	Market

Table 2.2: List of asset allocation models.
The table lists the considered asset allocation models along with its original (or prominent) reference. It is indicated if the model suffers from estimation risk in mean or variance. The last column states the abbreviation we use to refer to the strategy in the performance tables.

Data set (Source)	#Assets	Time Period	Investment category	Abbreviation
10 Industry Portfolios (Ken French's Web site)	10	07/1963-12/2011	Industry (low level)	Ind10
48 Industry Portfolios (Ken French's Web site)	48	07/1969-12/2011	Industry (high level)	Ind48
Dow Jones Industrial (Datastream)	23	02/1973-09/2012	Equities	Dow
3 Factor Fama French portfolio (MKT, SMB and HML) (Ken French's Web site)	3	07/1926-06/2012	Fama-French Factors	3Factor
Investment opportunities of Pension funds (Datastream)	4	01/1990-12/2010	Stocks, Real estate, Bonds, Commodities	Pension
World Market Portfolio (Datastream)	4	01/1988-09/2012	US, EU, Emerging Markets, Pacific	World

Table 2.3: List of data sets
The table lists the data sets considered. The first columns denote the name of the data set, its source, the number of assets n and the time period spanned. The investment category denotes the kind of asset the investor can diversify across. The last column states the abbreviation we use to refer to the data set in the performance tables. A description of the data sets is given in section 4.4.2.

Garlappi et al., 2007 or DeMiguel et al., 2009a). We use the following data sets: Data set *ind10* (*ind48*) consists of 10 (48, resp.) different industries in the US. The data set *Dow* replicates an investment in the Dow Jones Industrial firms. Unfortunately, at the beginning of the time series in 1973 seven companies were not founded yet. We restrict the data set to the 23 firms that already existed in 1973. Factor models such as the prominent Fama and French (1993) three factor (*3Factor*) model are included. To replicate the asset allocation of large *pension* funds, we consider the data set used by Hodder et al. (2012) consisting of stocks, bonds, real estate and commodities. To diversify over the whole *world*, we also include the portfolio of Weber et al. (2009) consisting of equities in North America, the Pacific Region, Europe and Emerging Markets as well as commodities. The empirical study should cover a convincing range of different investment opportunities which are of interest for large institutional investors as well as for private investors.

2.3.3 Evaluation

We carry out a comparison of the out-of-sample performance of the strategies presented in section 2.3.1 on the data sets introduced in section 4.4.2. The size of the rolling window is given by $h = 60$ months with robustness checks of $h = 120$ months[3]. For each time $t \in \{h+1, \dots, T\}$ and model i, the weights $\hat{w}_t^{(i)}$ are determined based on the past return vectors $\mathbf{R}_{t-h}, \dots, \mathbf{R}_{t-1}$ or the estimated moments $\hat{\mu}_t = \frac{1}{h}\sum_{s=1}^{h} \mathbf{R}_{t-s}$ and $\hat{\Sigma}_t = \frac{1}{h-1}\sum_{s=1}^{h}(\mathbf{R}_{t-s} - \hat{\mu}_t)^2$.The realized out-of-sample return of model i at time t is given by $r_t^{(i)} := \mathbf{R}_t' \hat{w}_t^{(i)}$. The performance of model i is determined based on the out-of-sample returns $\{r_t^{(i)}\}_{t=h+1}^{T}$. The moments are estimated by the corresponding sample moments. For example, the minimum variance of model i is given by $MinVar^{(i)} := -\frac{1}{T-h}\sum_{t=h+1}^{T}(r_t^{(i)} - \hat{\mu}_t^{(i)})^2$. Similar for the remaining performance functions stated in section 2.2.1. For the Certainty Equivalent a common value for the risk aversion is three, i.e. $\gamma = 3$ (Kan and Zhou, 2007). For robustness purpose we also analyzed a more risk-seeking investor ($\gamma = 1$) and a more risk-averse investor ($\gamma = 10$). As the obtained results are similar they are not stated.

The performance functions evaluate the overall performance of a certain model. To get an intuition how a model behaves at a specific time t, we provide plots in 2.C.1 presenting the change of the Certainty Equivalent over time. Other performance

[3] The insights from the results of $h = 120$ are similar to $h = 60$, thus we only report results for $h = 60$.

	CE	SR	Var	VaR	ES	Comb better (worse)	Average Position
Comb	0.071	0.234	0.1748	-0.0586	-0.0884	- (-)	2.5
	2.2 (1-3)	2 (1-3)	2.7 (2-3)	2.8 (2-4)	2.7 (2-3)		
MV (SR)	0.0751	0.2109	0.372	-0.0827	-0.1216	8 (1)	4.9
	2.8 (1-6)	4.2 (1-6)	6 (6-6)	5.8 (5-6)	5.8 (5-6)		
MinVar	0.0335	0.1689	0.1992	-0.0617	-0.0868	8 (3)	3.3
	5.2 (1-6)	4.5 (1-6)	2.2 (1-5)	2.5 (1-6)	2.2 (1-5)		
EQ	0.0681	0.2119	0.2349	-0.0643	-0.105	8 (0)	3.6
	3 (1-4)	3.3 (2-4)	4 (4-4)	3.3 (2-4)	4.2 (4-5)		
Market	0.0674	0.2061	0.2813	-0.0725	-0.1126	13 (0)	4.3
	3.3 (1-5)	3.8 (1-5)	4.7 (3-5)	4.8 (4-5)	4.7 (3-6)		
LW	0.0592	0.2237	0.132	-0.0525	-0.0752	2 (6)	2.5
	4.5 (2-5)	3.2 (1-5)	1.5 (1-2)	1.7 (1-3)	1.5 (1-2)		
Average	0.0607	0.2043	0.2439	-0.0667	-0.1003		

Table 2.4: Average Performance of Models

The average performance of a model with respect to a certain performance function is given. The number below represents the average ranking of the individual model over all data sets ($1 \triangleq$ best, $6 \triangleq$ worst). In brackets the best to worst position within the considered data sets is shown. The last two columns denote the frequence at which the combined model is at 10% significance level better (worse) than the individual model and average ranking over all data sets and performance functions.

functions perform similarly. The time specific CE is estimated based on the sample mean and variance of the past returns $r_{t-h}, \dots, r_{t-1}$ of the model considered, e.g.

$$\hat{CE}_t^{(i)} = \underbrace{\frac{1}{h}\sum_{s=1}^{h} r_{t-s}^{(i)}}_{\hat{\mu}_t^{(i)}} - \frac{\gamma}{2}\frac{1}{h-1}\sum_{s=1}^{h}(r_{t-s}^{(i)} - \hat{\mu}_t^{(i)})^2$$

2.3.4 Empirical Results

The detailed results in form of figures and tables are relegated to 2.C. The performance of all models with respect to the CE, Sharpe ratio, Variance, VaR and ES is presented. The main results are summarized in Table 2.4. The $p-$values are obtained by the methodology of Ledoit and Wolf (2008, 2011), which is summarized in 2.B.2.

The two main dimensions of the asset allocation problem are the applied data set and the chosen performance function. For a given performance function, we find that no model dominates in all data sets. For a given data set, we also find that no

	better 5% (worse)	better 10% (worse)	Comb better ind. Models	Comb better Average
CE	4 (0)	5 (0)	0.77	1
SR	8 (0)	10 (1)	0.80	0.75
Var	15 (8)	20 (9)	0.67	1
VaR	0 (0)	2 (0)	0.63	1
ES	0 (0)	2 (0)	0.67	1

Table 2.5: Comparison of individual models to the combined model
Comparison of the combined model to individual models and average. Given six data sets and five models, there are 30 direct comparisons of the combined model to individual models. The first two columns show how often the combined model is significantly better (worse) than an individual model at a 5% and 10% level. The third column shows the overall share of models which are worse than the combined model. The final column presents the share at which the combined model outperforms the average performance.

model dominates with respect to all performance functions. The best model is in any case influenced by the performance function and the underlying data. A obvious relationship between the ranking of models and the data set or performance function is not apparent. A model trained for a given performance function is not necessarily the out-of-sample best model (see e.g. variance of MinVar model). It is unlikely that the investor can ex-ante choose the best model for his applied performance function and his available investment universe. Additionally to the two dimensions, we find that the ranking of the models changes over time.

2.3.4.1 Individual Model Performance

For all models considered, Table 2.4 lists the performance in terms of different performance measures. The average performance along with the average ranking over all applied data sets is given. In brackets the best to worst position within the data sets is shown. First consider the average ranking as a measure of the performance of an individual model. Position (1) represents the best performance in all data sets, while the last position (6) denotes the worst performance. We find that the performance depends strongly on the performance function applied. The MVSR ranges from an intermediate position (2.8 for CE) to the worst position (5.8 for VaR, ES). The MinVar model performs well with respect to the variance (2.2) but bad with respect to the CE (5.2). The equally weighted and market weighted portfolio perform intermediated for all performance functions. The LW portfolio performs bad with respect to CE (4.5) but excellent with respect to Variance (1.5). We conclude that the performance of a model depends on the performance function applied.

Additionally, we find that the performance of the individual models depends strongly on the data sets considered. Consider the range of the best position to the worst position given in brackets in Table 2.4. For the CE the MVSR model is sometimes best (Dow) but can also be worst (Pension). The MinVar model ranges from best (3Factor) to worst (Ind10). The equally weighted appears to be the most stable one ranging from best (Pension) to forth best (Dow). Market can be best (Ind48) but also last but one (Pension). Finally, LW ranges from second (Pension) to last but one (Ind10). Similar results are obtained by the SR. For Var, VaR and ES the ranges are less broad. The MVSR portfolio performs less well, EQ performs intermediate while LW performs well. The performance of the MinVar model strongly varys from best to last but one. We note that the performance of a model does strongly depend on the data set applied.
It might be that the best model depends on the data set and the performance function applied but is at least stable over time. In order to answer this question, we consider the performance of the models over time for a given data set and a given performance function. The plots in section 2.C.1 show the performance of all models at each point in time. We present the figures for the CE only as the qualitative results are similar for other performance functions. We find that for a given performance function and data set there is no model which constantly outperforms the others.
We found that the performance of all models varies not only depending on the data set and performance function but also over time. We conclude that it is hard to determine the best model. Regarding this uncertainty, the combination of models seems to stabilize against the idiosyncratic risk of individual models.

2.3.4.2 Model Combination Performance

Now we turn to the empirical analysis of the combined model. Table 2.5 compares the performance of the combined model to the performance of the individual models. The final column of Table 2.5 shows the share over all data sets at which the performance of the combined model is better than the average performance of the individual models. For all performance functions but the Sharpe ratio, the performance of the combined model is constantly better than the average performance of all models. The share has to be one for the CE, Var and ES as these are concave performance functions. Still in case of the Sharpe ratio the performance of the combined model is better than the average portfolio in 75% of the cases. The combined model outperforms the average performance of the models. But what about direct comparisons?

The combined model can outperform the individual model but need not to. The combined model performs best if all models are equally well performing. In this case by Corollary 2.2.1, the combined model can beat all models. By the third column of Table 2.5, we find that the combined model is better than over 70% of all individual models. The first two columns of Table 2.5 show the number of cases at which the combined model significantly beats an individual model and the cases when the combined model is significantly beaten by an individual model. We consider the 5% and 10% significance level. It is over three times as likely that the combined model significantly beats an individual model than that it is significantly beaten by an individual model. For any performance function, the combined model outperforms more models significantly than it is outperformed by. The direct comparison shows that the combined model is rarely the best (only twice) but never the worst model. In Table 2.4 we find that the combined model is the second best portfolio in terms of the CE, Variance and VaR. It is even the best model in terms of the Sharpe ratio and the third best model for ES. The third but last line presents the frequency the combined model is significantly better than the individual model at a 10% significance level. The frequency the combined model is significantly worse than an individual model is given in brackets. Apart from the LW model we find that the combined model significantly outperforms all individual models more often than it is outperformed by. The naive strategies are never significantly better than the combined model. The only model which is better is the LW portfolio. We find that this portfolio performs in particular well with respect to the variance. The 6 situations at which the LW significantly outperforms the combined model are all with respect to this performance function. Considering the overall average ranking over all performance functions and data sets, we find that the combined model performs as well as the ex-post best model. The combined model and the LW have both an average rank of 2.5, while all other models are strongly worse.

We find that the combined model appears to be a very attractive strategy. It dominates most models with respect to all performance functions. We find that it performs almost as good as the ex-post best model and better than all remaining models. In the light of uncertainty of choosing the best model, it is doubtable if we had chosen ex-ante the ex-post best model. Model combination appears to be an attractive strategy. The strategy ensures ex-post to be above average - with the potential of being even the best model.

2.4 Conclusion

Portfolio optimization literature is characterized by the search of the best model. It is unlikely that there will ever be a single best model. The data generating process of returns is too complex and influenced by too many factors to be captured by any model. Some models perform well in certain times but fail in others. There is, however, no need to stick to one individual model. The simplest approach is to average across the models on hand. This approach has proven to work well in the context of forecasting. There are two conditions which lead to this finding: concavity of the performance function and changing model ranking. Most common performance functions in asset allocation are concave in standard settings. For concave performance functions the average model performs necessarily better than the average performance of the individual models. If the ranking of model changes in the course of time, the average portfolio constantly remains among the best models. Combination of models might but need not to be ex-post the best model. Ex-ante, the combined model is an attractive strategy as it is better than the average performance. An investor facing a set of models in expectation does necessarily better by taking the average model rather than randomly picking one model.

We run a large empirical analysis of six data sets with five models of three different model classes evaluated by five common performance functions. We find that no individual model dominates over a certain data set or performance function. We also show that the ranking of models changes over time. Therefore it is not surprising that the combination of models perform well. Apart from one case, the combined model is the only model which constantly stays in the best half of all models. Averaging over all data sets and performance functions the combination of models performs almost as well as the ex-post best individual model (Ledoit-Wolf) and outperforms all remaining models. Our results suggest that instead of relying on an individual model one should rather combine different models. So far we have only consider the naive equally weighting strategy for model combination. Future research should be based on finding more sophisticated combination schemes. Models performing well in the past should obtain a larger weight in the combined model.

Appendix 2.A Proofs

In order to ensure that the appendix is self-contained we repeat some of the comments of the paper. 2.A presents the proofs used in the paper. 2.B analyzes the CE performance function. Further we show how the $p-$values are obtained to compare the performance of two models. 2.C presents the empirical results. The overall results are displayed in tables. The performance of different models over time is shown by figures. The figures are given for the CE only as the qualitative performance is similar to the performance of other performance functions.

2.A.1 Proof of Lemma 2.2.1

LEMMA: For concave $F : \mathcal{A} \to \mathbb{R}$ it holds that

$$F\left(\sum_{i=1}^{m} \lambda_i x_i\right) \geq \sum_{i=1}^{m} \lambda_i F(x_i)$$

for weights Λ and $x_i \in \mathcal{A}$.

Proof. We show the Lemma by induction. The lemma holds for $n = 2$ as it is the definition of concavity. Assume it holds for m we show that it holds for $m + 1$ as well.

$$\begin{aligned} F\left(\sum_{i=1}^{m+1} \lambda_i x_i\right) &= F\left(\sum_{i=1}^{m} \lambda_i x_i + \lambda_m x_m\right) \\ &= F\left(\lambda x + (1-\lambda) x_m\right) \\ &\geq \lambda F\left(x\right) + (1-\lambda) F\left(x_m\right) \\ &= \lambda F\left(\sum_{i=1}^{m} \lambda_i x_i / \lambda\right) + \lambda_m F\left(x_m\right) \\ &\geq \sum_{i=1}^{m} \lambda_i F\left(x_i\right) + \lambda_m F\left(x_m\right) \\ &= \sum_{i=1}^{m+1} w_i F\left(x_i\right) \end{aligned}$$

where $x = \sum_{i=1}^{m} \lambda_i x_i / \lambda$ and $(1 - \lambda) = \lambda_m$. The last but one step made use of the induction's assumption. □

2.A.2 Concavity of CE

We show that $CE(\lambda r + (1-\lambda)\tilde{r}) \geq \lambda CE(r) + (1-\lambda)CE(\tilde{r})$.
Let the first two moments of r be μ and σ. Similar for $\tilde{r}$ with $\tilde{\mu}$ and $\tilde{\sigma}$. Then the CE of the combined model is given by

$$CE(\lambda r + (1-\lambda)\tilde{r}) = \lambda\mu + (1-\lambda)\tilde{\mu} - \frac{\gamma}{2}\left(\lambda^2\sigma^2 + (1-\lambda)^2\tilde{\sigma}^2 + 2\lambda(1-\lambda)\rho\sigma\tilde{\sigma}\right)$$

while the combined CE is given by

$$\lambda CE(r) + (1-\lambda)CE(\tilde{r}) = \lambda\mu - \lambda\frac{\gamma}{2}\sigma^2 + (1-\lambda)\tilde{\mu} - (1-\lambda)\frac{\gamma}{2}\tilde{\sigma}^2$$

Then the CE of the combined model is not smaller than the combined CE as

$$\begin{aligned}
CE(\lambda r + (1-\lambda)\tilde{r}) &\geq \lambda CE(r) + (1-\lambda)CE(\tilde{r}) \\
-\frac{\gamma}{2}\left(\lambda^2\sigma^2 + (1-\lambda)^2\tilde{\sigma}^2 + 2\lambda(1-\lambda)\rho\sigma\tilde{\sigma}\right) &\geq -\lambda\frac{\gamma}{2}\sigma^2 - (1-\lambda)\frac{\gamma}{2}\tilde{\sigma}^2 \\
\lambda(1-\lambda)\sigma^2 + \lambda(1-\lambda)\tilde{\sigma}^2 - 2\lambda(1-\lambda)\rho\sigma\tilde{\sigma} &\geq 0 \\
\sigma^2 + \tilde{\sigma}^2 - 2\rho\sigma\tilde{\sigma} &\geq 0
\end{aligned}$$

The last equation holds as $\sigma^2 + \tilde{\sigma}^2 - 2\rho\sigma\tilde{\sigma} \geq (\sigma - \tilde{\sigma})^2 \geq 0$. Therefore not that the equality holds strict unless $\rho = 1$ and $\sigma^2 = \tilde{\sigma}^2$. If the assets with the same risk are perfectly correlated there is no improvement of the combination.

2.A.3 (Non-)Concavity of Sharpe Ratio

We derive the conditions for concavity and dominance in average returns of the Sharpe ratio. The Sharpe ratio of the combined model is given by

$$\begin{aligned}
\frac{\sum_i w_i\mu_i}{\sqrt{V\left(\sum_i w_i R_i\right)}} &= \frac{\sum_i w_i\mu_i}{\sqrt{\sum_i w_i w_j \sigma_i\sigma_j\rho_{ij}}} \\
&\geq \frac{\sum_i w_i\mu_i}{\sqrt{\left(\sum_i w_i\sigma_i\right)^2}} = \frac{\sum_i w_i\mu_i}{\sum_i w_i\sigma_i}
\end{aligned}$$

Now we compare the lower bound of the combined model with the combined Sharpe ratio of the individual models

$$\begin{aligned}
\sum_i w_i SR_i &\leq SR_{Comb} \\
\sum_i w_i \frac{\mu_i}{\sigma_i} &\leq \frac{\sum_i w_i \mu_i}{\sum_i w_i \sigma_i} \\
\sum_i w_i \mu_i (w_i + \sum_{j \neq i} w_j \frac{\sigma_j}{\sigma_i}) &\leq \sum w_i \mu_i \\
0 &\leq \sum_i w_i \mu_i \left(1 - w_i - \sum_{j \neq i} w_j \frac{\sigma_j}{\sigma_i}\right) \\
0 &\leq \sum_i w_i SR_i \left((1 - w_i)\sigma_i - \sum_{j \neq i} w_j \sigma_j\right) \\
0 &\leq \sum_i w_i SR_i \left(\sum_{j \neq i} w_j (\sigma_i - \sigma_j)\right)
\end{aligned}$$

The non-negativity of $\sum_i w_i SR_i \left(\sum_{j \neq i} w_j (\sigma_i - \sigma_j)\right)$ is a sufficient condition for the Sharpe Ratio to be concave. We consider now two special cases:
1.) Let $w_j = \frac{1}{n}$ for all $j \in \{1, \ldots, n\}$. Then the condition reduces to $\sum_i SR_i \left(\sum_{j \neq i} (\sigma_i - \sigma_j)\right)$ being non-negative. The condition is satisfied if the risky models have a higher Sharpe ratio. The degree the Sharpe ratio has to be higher depends on weights w.
2.) Consider $n = 2$ with $w_1 = 1 - w_2 = \lambda$. Then the condition reduces to

$$\begin{aligned}
0 &\leq \lambda SR_1 (1 - \lambda)(\sigma_1 - \sigma_2) \\
&\quad + (1 - \lambda) SR_2 \lambda (\sigma_2 - \sigma_1) \\
0 &\leq (SR_1 - SR_2)(\sigma_1 - \sigma_2)
\end{aligned}$$

which is satisfied if $SR_1 \geq SR_2$ and $\sigma_1^2 \geq \sigma_2^2$ or $SR_1 \leq SR_2$ and $\sigma_1^2 \leq \sigma_2^2$. Again the condition is that the risky model has a higher Sharpe ratio.
However, the dominance in average returns of the Sharpe ratio does not hold generally. As a counterexample let $\mu = \tilde{\mu} = 0.1$ and $\sigma = 0.1$, $\tilde{\sigma} = 0.2$. The two returns are strongly correlated ($\rho = 1$). Then the equally weighted Sharpe ratio is given by $\frac{1}{2} SR(r) + \frac{1}{2} SR(\tilde{r}) = \frac{3}{4}$. The Sharpe ratio of the equally weighted returns is only $SR\left(\frac{r + \tilde{r}}{2}\right) = \frac{2}{3}$. Hence, in expectation it is better not to combine but to randomly select an individual asset.

2.A.4 (Non-)Concavity of VaR

The value at risk is a concave risk preference if the returns' VaR can be represented by $VaR_\alpha(R_i) = \mu_i - z_\alpha \sigma_i$. That assumption is satisfied for many distributions such as the $t-$distribution. Then it holds that

$$\begin{aligned}
VaR\left(\sum_i w_i r_i\right) &= \sum_i w_i \mu_i - z_\alpha \sqrt{V(\sum_i w_i R_i)} \\
&= \sum_i w_i \mu_i - z_\alpha \sqrt{\sum_i w_i^2 \sigma_i^2 + \sum_{i \neq j} w_i w_j \sigma_i \sigma_j \rho_{ij}} \\
&\geq \sum_i w_i \mu_i - z_\alpha \sqrt{\left(\sum_i w_i \sigma_i\right)^2} \\
&= \sum_i w_i \mu_i - z_\alpha \sum_i w_i \sigma_i \\
&= \sum_i w_i \mu_i - z_\alpha w_i \sigma_i \\
&= \sum_i w_i VaR(r_i)
\end{aligned}$$

Under the condition that the VaR can be linearly calculated of the first two moments we find that the VaR is a concave performance function.

It can be easily seen by the following counter-example that this does not hold generally. Consider two returns $R, \tilde{R}$ both are independent binary distributed with $Pr(R = 1) = Pr(\tilde{R} = 1) = 0.96$. Then the equally weighted return is given by

$$\frac{R + \tilde{R}}{2} = \begin{cases} 0 & p = 0.04^2 \\ 0.5 & p = 2 \cdot 0.04 \cdot 0.96 \\ 1 & p = 0.96^2 \end{cases}$$

With significance level $\alpha = 0.05$ it follows that

$$VaR\left(\frac{R + \tilde{R}}{2}\right) = 0.5 < \frac{1}{2}(1 + 1) = \frac{1}{2}\left(VaR(R) + VaR(\tilde{R})\right)$$

2.A.5 Concavity of ES

Using the dual representation of Foellmer and Schied (2010) we can write

$$ES_\alpha(R) = \inf_{Q \in Q_\alpha} E_Q(R)$$

with Q_α is a set of probability measures which are absolutely continuous to the physical measure P such that the Radon-Nikodym derivative of Q with respect to P is almost surely bounded by $1/\alpha$, i.e. $\frac{dQ}{dP} \leq \alpha^{-1}$ a.s.
Then it holds that

$$\begin{aligned} ES\left[\lambda R + (1-\lambda)\tilde{R}\right] &= \inf_{Q \in Q_\alpha} E_Q\left[\lambda R + (1-\lambda)\tilde{R}\right] \\ &= \inf_{Q \in Q_\alpha} \left(\lambda E_Q\left[R\right] + (1-\lambda) E_Q\left[\tilde{R}\right]\right) \\ &\geq \lambda \inf_{Q \in Q_\alpha} E_Q\left[R\right] + (1-\lambda) \inf_{\tilde{Q} \in Q_\alpha} E_{\tilde{Q}}\left[\tilde{R}\right] \\ &= \lambda ES\left[R\right] + (1-\lambda) ES\left[\tilde{R}\right] \end{aligned}$$

Appendix 2.B Performance Functions

2.B.1 Certainty Equivalent

We present situations when the utility function approximately corresponds to the CE (eq. 2.3). The argumentation is similar to Frahm (2012). Let r be the assets' returns, $r_p = w_p' r$ be the return of the portfolio and the initial wealth of the investor be denoted by $v > 0$.
1.) If U is quadratic. This point is trivial.
2.) If r is normal distributed and the investor has a exponential utility function U. As he is invested completely his wealth changes to $v + vr_p$. It holds that

$$\begin{aligned} EU(v + vr_p) &= E\left[-\exp\left(-\gamma(v + vr_p)\right)\right] \\ &= -\exp\left(-\gamma v\left[1 + E(r_p) - \frac{\gamma v}{2} V(r_p)\right]\right) \\ &\sim E(r_p) - \frac{\tilde{\gamma}}{2} V(r_p) \end{aligned}$$

with risk aversion parameter $\tilde{\gamma} = \gamma v$. The expected utility is maximized for maximizing equation 2.3.
3.) If investment horizon h is short. Let the utility function U and return process r

be nice[4]. The investor's return of his portfolio w is given by $r_p = w_p' rh$. His utility can be represented by the following Taylor series

$$\begin{aligned} U(v + v \cdot r_p) &= U(v) + U'(v) v \cdot r_p + \frac{1}{2} U''(v) \left(v \cdot r_p\right)^2 + \frac{1}{6} U''' \left(\xi\right)^3 \\ &\overset{h \to 0}{\longrightarrow} U(v) + U'(v) \cdot v \cdot r_p + \frac{1}{2} U''(v) \left(v \cdot r_p\right)^2 \\ &= U(v) + U'(v) v \left[r_p - \frac{1}{2} \left(-\frac{U''(v)}{U'(v)} v \right) r_p^2 \right] \end{aligned}$$

where $\xi \in [v, v + v r_p]$. For a risk averse investor

$$\alpha = -\frac{U''(v)}{U'(v)} v > 0$$

denotes the Arrow-Pratt measure of relative risk aversion (see Arrow, 1971; Pratt, 1964). Then the expected utility is approximately given by

$$\begin{aligned} EU(v + v \cdot r_p) &\approx EU(v) + U'(v) v \left[E\left(r_p\right) - \frac{\alpha}{2} E\left(r_p^2\right) \right] \\ &\approx EU(v) + U'(v) v \left[E\left(r_p\right) - \frac{\alpha}{2} V\left(r_p\right) \right] \\ &\sim E\left(r_p\right) - \frac{\alpha}{2} V\left(r_p\right) \end{aligned}$$

Hence the expected utility for short term horizon approximately corresponds to the objective function given in eq. 2.3.

2.B.2 p-values

To test if a strategy i is significantly different to the combination strategy c, we derive the $p-$values. For CE, Sharpe ratio and variance we use the bootstrapping methodology proposed in Ledoit and Wolf (2008, 2011). Their approach accounts for common characteristics of portfolio returns such as fat tails or time series dependence, such as serially correlation or volatility clustering. The circular block bootstrap of Politis and Romano (1994) is used. As in Ledoit and Wolf (2008) or DeMiguel et al. (2009a) the block size is chosen to be equal to $b = 5$ with $M = 1000$ bootstrap resamples. We find that changes to block size or resampling size does not significantly alter the results. The vector of moments of strategy i and combination c is given by μ_i, σ_i^2 and μ_c, σ_c^2, resp. Let the vector of moments be $\eta = (\mu_i, \mu_c, \sigma_i^2, \sigma_c^2)$

[4] Let the utility function U be three times continuous differentiable and the return process r follow a multivariate Brownian motion.

and $\hat{\eta}$ be its empirical counterpart. The difference of CE, Sharpe ratio and variance is given by

$$f_{CE}(\eta) = \mu_i - \frac{\gamma}{2}\sigma_i^2 - \mu_c + \frac{\gamma}{2}\sigma_c^2$$

$$f_{SR}(\eta) = \frac{\mu_i}{\sigma_i} - \frac{\mu_c}{\sigma_c}$$

$$f_{Var}(\eta) = \log \sigma_i^2 - \log \sigma_c^2$$

The corresponding sample value is denoted by $\hat{f}$. We assume for the moments that it holds (see Press, 1972, p. 107) that

$$\sqrt{T}(\eta - \hat{\eta}) \xrightarrow{d} N(0, \Psi)$$

Under certain regularization conditions (see Greene, 2002) the delta method implies that

$$\sqrt{T}(f - \hat{f}) \xrightarrow{d} N\left(0, \nabla' f(\eta)\Psi\nabla f(\eta)\right)$$

with ∇f being the derivative of f. We estimate Ψ by kernal estimation using the prewhitened QS kernel of Andrews and Monahan (1992). The standard errors

$$s(\hat{f}^{\star,m}) = \sqrt{\frac{\nabla' \hat{f}^{\star,m}\hat{\Psi}\nabla\hat{f}^{\star,m}}{T}}$$

are calculated for each boostrap repetition $m = 1, \ldots, M$. Then the p-value is given by

$$PV = \left(\# \left\{ \frac{\left|\hat{f}^{\star,m} - \hat{f}\right|}{s(\hat{f}^{\star,m})} \geq \frac{\left|\hat{f}\right|}{s(\hat{f})} \right\} + 1 \right) / (M+1)$$

For the Sharpe ratio and the variance we use the code available at http://www.econ.uzh.ch/faculty/wolf . The modification for the CE is straight forward.

The p-values for VaR and ES are obtained by a circular block bootstrap with $b = 5$.

Appendix 2.C Empirical Study

2.C.1 CE over time

The following figures show the estimated performance of all models over time.

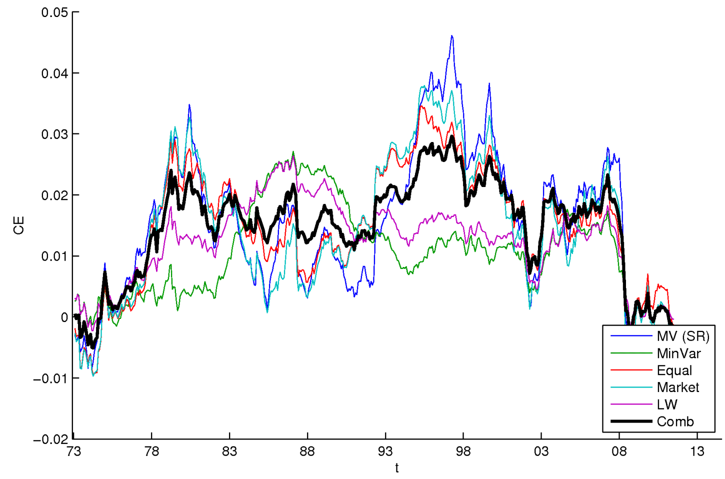


Figure 2.C.3: Ind10

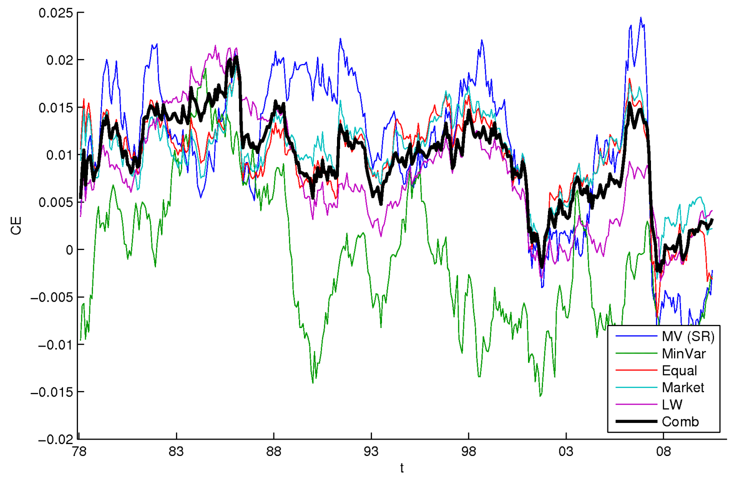


Figure 2.C.4: Ind48

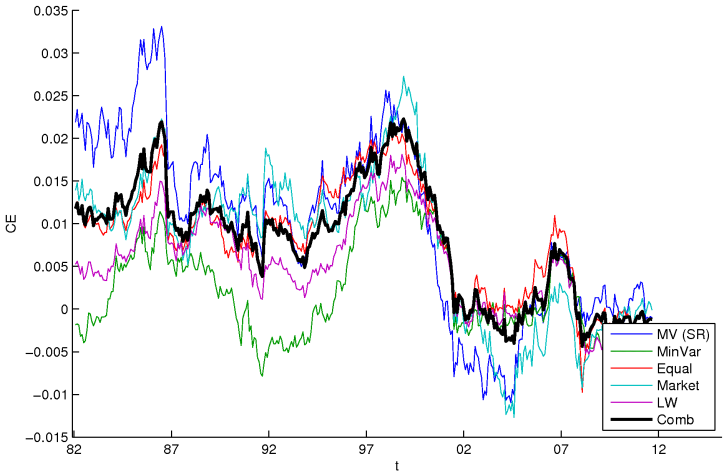


Figure 2.C.5: Dow

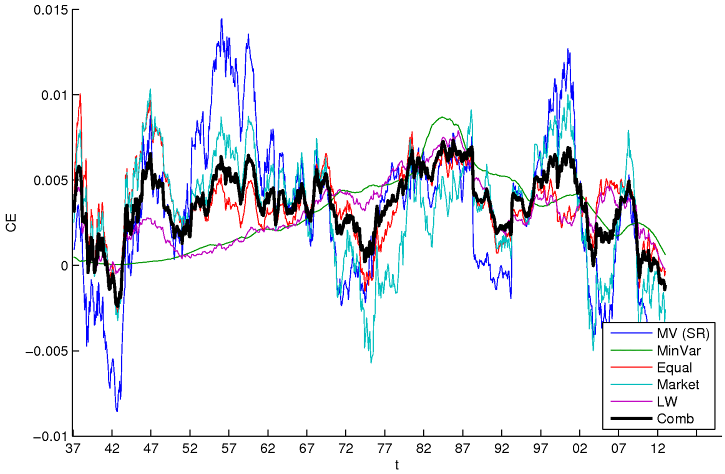


Figure 2.C.6: 3Factor

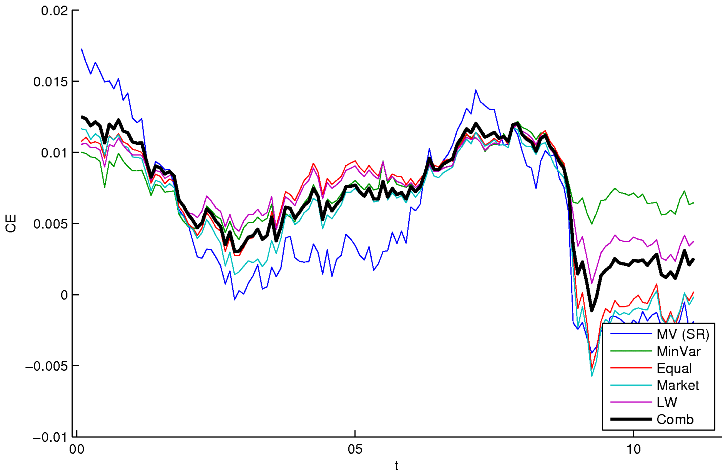


Figure 2.C.7: Pension

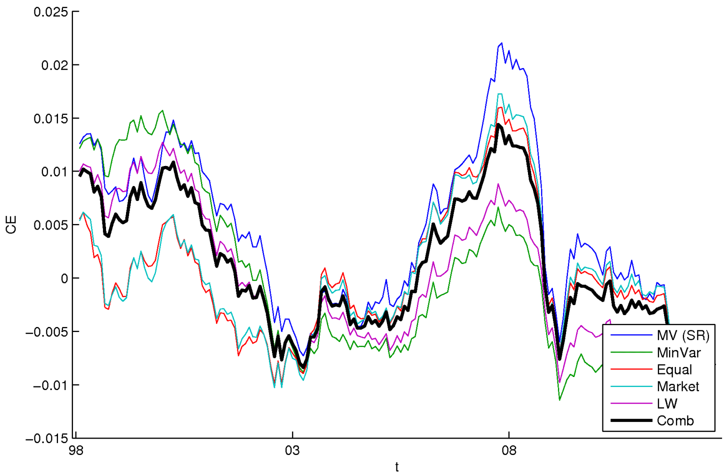


Figure 2.C.8: World

2.C.2 Performance

The performance of each model in each data set with respect to the five performance functions is presented.

	Ind10	Ind48	Dow	3 Factor	Pension	World
Comb	0.1319	0.0867	0.0791	0.0315	0.0752	0.0218
MVSR	0.1257	0.0903	0.1016	0.0286	0.0612	0.0431
	(0.3811)	(0.5514)	(0.8856)	(0.3357)	(0.2363)	(0.8322)
MinVar	0.0910	-0.0184	0.0185	0.0254	0.0812	0.0034
	(0.045)	(0.001)	(0.002)	(0.0764)	(0.6184)	(0.1499)
EQ	0.1312	0.0859	0.0769	0.0322	0.0686	0.0140
	(0.4775)	(0.482)	(0.4221)	(0.6034)	(0.3452)	(0.2113)
Market	0.1253	0.0937	0.0786	0.0297	0.0634	0.0136
	(0.3477)	(0.7403)	(0.488)	(0.3611)	(0.2203)	(0.2263)
LW	0.1155	0.0743	0.0498	0.0279	0.0774	0.0104
	(0.1748)	(0.2033)	(0.0045)	(0.1693)	(0.5849)	(0.1568)
Av. Perf	0.1177	0.0652	0.0651	0.0288	0.0703	0.0169

Table 2.C.6: CE

	Ind10	Ind48	Dow	3 Factor	Pension	World
Comb	0.3426	0.2649	0.2441	0.1561	0.2784	0.1179
MVSR	0.2821	0.2335	0.2507	0.1333	0.2032	0.1624
	(0.012)	(0.1868)	(0.5944)	(0.1144)	(0.035)	(0.9461)
MinVar	0.3011	0.0787	0.1099	0.1386	0.3117	0.0732
	(0.2168)	(0.001)	(0.001)	(0.1499)	(0.6479)	(0.1044)
EQ	0.3050	0.2431	0.2295	0.1542	0.2344	0.1053
	(0.0165)	(0.2173)	(0.2932)	(0.4171)	(0.0959)	(0.2867)
Market	0.2839	0.2653	0.2208	0.1372	0.2205	0.1087
	(0.001)	(0.5065)	(0.1713)	(0.0864)	(0.0435)	(0.2987)
LW	0.3702	0.2545	0.1838	0.1483	0.2957	0.0895
	(0.7707)	(0.3991)	(0.016)	(0.2732)	(0.6653)	(0.1044)
Av. Perf	0.3085	0.2150	0.1989	0.1423	0.2531	0.1078

Table 2.C.7: Sharpe

	Ind10	Ind48	Dow	3 Factor	Pension	World
Comb	0.2402	0.1880	0.1999	0.0750	0.1076	0.2379
MVSR	0.5268	0.5167	0.4761	0.1293	0.2040	0.3792
	(0.0492)	(0.0397)	(0.0397)	(0.056)	(0.0778)	(0.0359)
MinVar	0.1373	0.4901	0.2214	0.0635	0.0932	0.1894
	(0.9548)	(0.06)	(0.0419)	(0.9574)	(0.8278)	(0.9566)
EQ	0.3820	0.2710	0.2456	0.0848	0.1521	0.2740
	(0.0224)	(0.0354)	(0.0396)	(0.0263)	(0.0632)	(0.0154)
Market	0.4919	0.2376	0.3620	0.1246	0.1538	0.3176
	(0.0242)	(0.0277)	(0.0301)	(0.0376)	(0.072)	(0.0172)
LW	0.1341	0.1407	0.1638	0.0642	0.0965	0.1928
	(0.9699)	(0.9567)	(0.9709)	(0.9693)	(0.9092)	(0.9729)
Av. Perf	0.3344	0.3312	0.2938	0.0933	0.1399	0.2706

Table 2.C.8: Var

	Ind10	Ind48	Dow	3 Factor	Pension	World
Comb	-0.0594	-0.0632	-0.0652	-0.0333	-0.0461	-0.0845
MVSR	-0.0913	-0.0993	-0.1008	-0.0516	-0.0564	-0.0969
	(0.1316)	(0.1526)	(0.1276)	(0.1515)	(0.1998)	(0.2855)
MinVar	-0.0492	-0.1049	-0.0745	-0.0281	-0.0407	-0.0730
	(0.7058)	(0.1318)	(0.3846)	(0.7463)	(0.6191)	(0.6154)
EQ	-0.0796	-0.0697	-0.0665	-0.0360	-0.0434	-0.0903
	(0.1066)	(0.3725)	(0.3442)	(0.4363)	(0.4673)	(0.2953)
Market	-0.0877	-0.0724	-0.0803	-0.0511	-0.0490	-0.0943
	(0.0545)	(0.3614)	(0.2199)	(0.0739)	(0.2622)	(0.1925)
LW	-0.0464	-0.0526	-0.0620	-0.0289	-0.0444	-0.0805
	(0.8226)	(0.7008)	(0.6318)	(0.7795)	(0.7369)	(0.7191)
Av. Perf	-0.0708	-0.0798	-0.0768	-0.0391	-0.0468	-0.0870

Table 2.C.9: VaR

	Ind10	Ind48	Dow	3 Factor	Pension	World
Comb	-0.1029	-0.0970	-0.0914	-0.0546	-0.0716	-0.1130
MVSR	-0.1439	-0.1512	-0.1268	-0.0790	-0.0998	-0.1291
	(0.1334)	(0.1585)	(0.1183)	(0.1484)	(0.2041)	(0.2961)
MinVar	-0.0720	-0.1510	-0.0995	-0.0416	-0.0506	-0.1058
	(0.6745)	(0.1279)	(0.38)	(0.7353)	(0.5735)	(0.5773)
EQ	-0.1265	-0.1193	-0.1095	-0.0594	-0.0967	-0.1189
	(0.1162)	(0.3575)	(0.34)	(0.4114)	(0.4632)	(0.2948)
Market	-0.1455	-0.1081	-0.1229	-0.0758	-0.0954	-0.1281
	(0.0511)	(0.3831)	(0.2275)	(0.0897)	(0.2555)	(0.2092)
LW	-0.0737	-0.0841	-0.0834	-0.0451	-0.0625	-0.1026
	(0.8024)	(0.682)	(0.6377)	(0.766)	(0.7077)	(0.6667)
Av. Perf	-0.1123	-0.1227	-0.1084	-0.0602	-0.0810	-0.1169

Table 2.C.10: ES

Bibliography

ANDREWS, D. AND J. MONAHAN (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953–966.

ARROW, K. (1971): *Aspects of the Theory of Risk Bearing*, Markham, chap. The theory of risk aversion, 90–109.

ARTZNER, P., F. DELBAEN, J. EBER, AND D. HEATH (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9 (3), 203–228.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1997): "Thinking coherently," *RISK*, 10 (11), 68–71.

BATES, J. AND C. GRANGER (1969): "The combination of forecasts," *Operational Research Quarterly*, 20, 451–468.

BEST, M. AND R. GRAUER (1991): "On the sensitivity of mean variance efficient portfolios to changes in asset means: some analytical and computational results," *Review of Financial Studies*, 4 (2), 315–342.

BODIE, Z., A. KANE, AND A. J. MARCUS (1993): *Investments*, Homewood.

BRITTEN-JONES, M. (1999): "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *Journal of Finance*, 54(2), 655–671.

CHAN, L. K. C., J. KARCESKI, AND J. LAKONISHOK (1999): "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model," *Review of Financial Studies*, 12, 937–974.

CHO, D. (2011): "Estimation risk in covariance," *Journal of Asset Management*, 12, 248–259.

CLEMEN, R. (1989): "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, 5, 559–581.

DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): "Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms." *Management Science*, 55(5), 798–812.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): "Optimal Versus Naive

Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22 (5), 1915–1953.

Diebold, F. and P. Pauly (1987): "Structural change and the combination of forecasts," *Journal of Forecasting*, 6, 21–40.

Embrechts, P., A. J. McNeil, and D. Straumann (2002): *Correlation and Dependence in Risk Management: Properties and Pitfalls*, Cambridge University Press.

Fama, E. F. and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1), 3–56.

Fan, J., J. Zhang, and K. Yu (2012): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios," *Journal of American Statistical Association*, 107, 592–606.

Foellmer, H. and A. Schied (2002): "Convex measures of risk and trading constraints," *Finance and Stochastics*, 6 (4), 429–447.

——— (2010): *Encyclopedia of Quantitative Finance*, John Wiley & Sons, chap. Coherent and convex risk measures, 355–363.

Frahm, G. (2012): "Capital Allocation Under Risk and Uncertainty," *Working Paper*.

Frost, P. A. and J. E. Savarino (1986): "An empirical Bayes approach to portfolio selection." *Journal of Financial and Quantitative Analysis*, 21, 293–305.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *The Review of Financial Studies*, 20, 41–81.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013): "Combining Expert Forecasts: Can Anything Beat the Simple Average?" *International Journal of Forecasting*.

Greene, W. H. (2002): *Econometric Analysis*, Prentice-Hall.

Hodder, J., J. Jackwerth, and O. Kolokolova (2012): "Improved Portfolio Choice Using Second-Order Stochastic Dominance," *Working Paper*.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps." *Journal of Finance*, 54(4), 1651–1684.

Jorion, P. (1986): "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21(3), 2279–292.

——— (1991): "Bayesian and CAPM estimators of the means: Implications for portfolio selection." *Journal of Banking and Finance*, 15, 717–727.

Kan, R. and G. Zhou (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *Journal of Financial and Quantitative Analysis*, 42(3), 621–656.

Krueger, F. and P. Schanbacher (2012): "Strictly Proper Scoring Rules and the Wisdom of the Crowd," *University of Konstanz, Working Paper.*

Ledoit, O. and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004a): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 21(1), 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2011): "Robust performance hypothesis testing with the variance," *Wilmott Magazine*, 86–89.

Lichtendahl, K. C., Y. Grushka-Cockayne, and R. L. Winkler (2012): "Is it Better to Average Probabilities or Quantiles?" *Management Science (forthcoming).*

Markowitz, H. M. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

Merton, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics*, 8, 323–361.

Politis, D. and J. Romano (1994): "The stationary bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Pratt, J. (1964): "Risk aversion in the small and in the large," *Econometrica*, 32, 122–136.

Press, J. (1972): *Applied Multivariate Analysis*, Holt, Rinehart and Winston, New York.

Radcliff, R. C. (1990): *Investment Concepts, Analysis, Strategy*, HarperCollins.

Sharpe, W. F. (1963): "A simplifed model for portfolio analysis." *Management Science*, 9, 277–293.

——— (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance*, 19(3), 425–442.

——— (1994): "The Sharpe Ratio," *The Journal of Portfolio Management*, 21 (1), 49–58.

Smith, J. and K. Wallis (2009): "A simple explanation of the forecast combination puzzle," *Oxford Bulletin of Economics and Statistics*, 71 (3), 331–355.

Stock, J. and M. Watson (2004): "Combination forecasts of output growth in a seven-country data set," *Journal of Forecasting*, 23, 405–430.

Surowiecki, J. (2004): *The Wisdom of Crowds.*, RandomHouse, NewYork.

Tu, J. and G. Zhou (2011): "Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies," *Journal of Financial Economics*, 99, 204–215.

Varian, H. R. (1992): *Microeconomic Analysis (Third ed.).*, W.W. Norton and Company.

Weber, M., H. Jacobs, and S. Müller (2009): "How Should Private Investors Diversify? - An Empirical Evaluation of Alternative Asset Allocation Policies to Construct A World Market Portfolio," *Working Paper*.

Chapter 3

MINIMAX: Portfolio Choice Based on Pessimistic Decision Making

3.1 Introduction

Taking risk is usually accompanied by obtaining economic gains. In portfolio management, an investor typically trades risk against profits, according to his preference. However, in this trade-off it is fundamental to define "risk" appropriately. In the typical Mean-Variance model of Markowitz (1952), risk is defined as variation in returns (for a discussion of different risk measures, see Biglova et al., 2004). In this paper, we deviate from classical risk measures and define risk in terms of a worst case scenario for two reasons. First, symmetric risk measures incorporate undesirable properties. Positive deviations of returns from their means should not be considered as risk, as volatility implies. Second, considering risk as measured by worst case scenarios is reasonable given recent periods of economic turmoil. During periods of moderate economic changes, risk measures such as volatility or Value-at-Risk may be quite effective. In extreme economic periods, an investor may prefer to be much more conservative. For both reasons, we consider the highest realized loss to be the risk that matters most for an investor and consequently, we apply this risk measure for portfolio optimization.

Turning to first moments of return distributions, it has been shown in the past that estimating sample means suffers from serious problems. Slightly different mean estimates of asset returns can result in completely adverse portfolio positions. The portfolio literature proposes different ways to handle this estimation error (see e.g. Jorion, 1986). Similar problems occur when estimating the covariance matrix (Ledoit and Wolf, 2003). Minimax optimization not only implements a more appropriate risk measure for risk averse investors, but also avoids the major problem of estimating means and covariances.

A typical investor that may apply the Minimax trading strategy is a corporation, a pension fund, or a bank that invests in several asset classes and that is facing high regulatory requirements such as Basel II and Basel III. This regulation often requires daily risk management. Consequently, the Minimax strategy meets the needs of institutional investors that aim at minimizing daily investment losses.

The second type of investor that needs to minimize daily investment losses is an investor who is facing mark-to-market accounting, either for regulatory issues or, e.g., because of leveraged investment positions that may trigger margin calls if the portfolio falls short of a certain level.

In order to find an "optimal portfolio" that satisfies any risk-averse investor, different portfolio selection rules have been proposed by a large body of finance research. To be mentioned first, the naive 1/N allocation rule goes back to the fourth century[1]and survived until today. With this simple rule of thumb, one can obtain a well-diversified and risk reducing portfolio without having to estimate any model parameter. It is still a plausible and "hard to beat" benchmark for newly proposed portfolio selection methods. DeMiguel et al. (2009b) state and show that there are still "miles to go" to beat the 1/N strategy, in particular for sample-based Mean-Variance strategies. More and more of sophisticated asset allocation rules have been proposed over time. Markowitz (1952) derived the Mean-Variance theory, and still today, portfolio optimization based on the first two moments of the return is widely used. However, it suffers from serious caveats. First, it assumes either that the investor has an approximately quadratic utility function or that returns that are normally distributed and the investor's utility function is exponential. Second, estimating moments via their sample analogues leads to extreme and fluctuating weights[2], and as a result, portfolios perform poorly out-of-sample, in particular when trading costs are considered or when return distributions deviate from normality; see e.g. Michaud (1989). A large literature examines extensions of the popular Mean-Variance model to reduce estimation error, for instance the Bayesian approach. In such a framework, Kan and Zhou (2007) combine Mean-Variance, Minimum Variance portfolios as well as the risk-free asset in order to decrease the influence of estimation error on investor's utility. Additionally, DeMiguel et al. (2009b) show that imposing L_2 norm constraints is equivalent to shrinking the covariance matrix. Lasso constraints are analyzed by Fan et al. (2012). For a summary of extensions to the Mean-Variance model, see DeMiguel et al. (2009a). We want to propose a trading strategy that fits any highly risk-averse investor, and to the best of our knowledge, we are the first to examine the performance of the Minimax strategy in a multi-asset-class portfolio choice context. The Minimax strategy is a "pessimistic" trading strategy, because it chooses portfolio weights such that the portfolio payoff is maximized in the worst case scenario. We use one year of daily returns. Different portfolio weights lead to different daily returns. We consider the worst realized daily return in the one year period only. The Minimax portfolio is given by the portfolio which leads to the highest return (i.e. smallest loss) in the worst case situation.

[1] "One should always divide his wealth into three parts: a third in land, a third in merchandise, and a third ready to hand." This basic asset allocation rule was already proposed by Rabbi Issac bar Aha. See Babylonian Talmud: Tractate Baba Mezi'a, folio 42a, 4 century AD.

[2] Merton (1980) shows that sample estimates of expected returns are very unstable. As a result, standard errors of portfolio weights are huge; see, e.g. Britten-Jones (1999). Michaud (1989) even considers Mean-Variance optimizations as a maximization of errors.

The first one that introduces the term “Minimax” in a portfolio context is Young (1998). He suggests linear programming for maximizing the minimum return of a portfolio based on historical returns, given a certain required minimum return. There have been earlier studies by, e.g., Sengupta (1982) and Lintner and Krasker (1982) on minimax optimization in portfolio choice. By a simulation study, Young (1998) is the first to find that the Minimax model outperforms traditional Mean-Variance portfolios. Further, he showed that the Minimax model is compatible with expected utility maximization. There is a large body of research on linear programming models such as the Minimax model, and Mansini et al. (2003) provide a systematic overview as well as a discussion of their properties. Ding (2006) considers Minimax models based on quantiles rather than the minimum return. Analytical results in the context of portfolio optimization are provided by Cai et al. (2000) and Teo and Yang (2001).

This literature also connects Minimax modeling to ambiguity aversion, another branch of finance research. Gilboa and Schmeidler (1989) model situations in which a decision maker has not enough information to assume one single prior distribution and determines investor preference as utility function over a set of multiple prior distributions. Chateauneuf et al. (2005) derive the theoretical framework for some important applications of multiple priors. In a recent paper, Garlappi et al. (2007) use confidence intervals for expected returns to model decision-making under multiple priors. They also show that if ambiguity aversion goes to infinity, the resulting portfolio is the global Minimum Variance portfolio. We will see later that the Minimum Variance portfolio is the hardest benchmark to beat in terms of portfolio risk and portfolio performance. The closest study to ours is Tuetuencue and Koenig (2004). They use uncertainty sets for the moments of returns to obtain portfolios with best worst case behavior. Still, in their numerical exercise, they only examine equity and fixed income securities. We contribute to the literature by showing empirically that the Minimax portfolio choice strategies are valid and practically implementable for a representative multi asset investor, and we compare the performance of our Minimax trading strategy to common alternative benchmarks with respect to different performance measures.

We use US stock, bond, real estate and commodity indexes to construct portfolios with yearly portfolio holding periods. Our representative investor is a multi asset investor who is facing stringent risk reporting - such as a large pension fund - and thus, cares about daily investment losses. The largest stake of US pension funds is invested in stocks and bonds. Yet, many funds diversify into real estate, commodities and other alternative investments as well. Looking at the Public Fund Survey 2009 of Brainard, we find that a representative US pension fund invested about 60% of its assets into stocks, about 20% into bonds, and the remaining part in other assets, such as real estate, private equity, hedge funds and commodities. Additionally, Belousova and Dorfleitner (2012) find that commodities add significant diversification potential to a portfolio of European stocks and bonds, and commodities can be particularly beneficial for risk-minimizing portfolio strategies. In order to compare our results to the US pension fund industry, we build a fixed weight benchmark that mimics a constant investment style of a representative US pension fund. Of course, we also test our results against the naive diversification strategy, which invests one fourth of funds into each asset class. In addition, we construct benchmarks that are based on common alternative portfolio selection concepts: Mean-Variance and Minimum Variance optimization.

In our empirical analysis, we find that the Minimax portfolio outperforms the considered benchmarks. From a practical perspective, it is important to note that the weights of our Minimax trading strategy are stable over time and consequently, our Minimax strategy has a lower turnover than all Mean-Variance related strategies, even in the presence of short-selling constraints. This remark is noteworthy, because Minimax optimization leads to positive weights per definition, which is an appreciable feature from the viewpoint of a practitioner. Furthermore, the resulting Minimax portfolios are diversified across all considered asset classes with on average 11% invested in stocks, 54% invested in bonds, 15% invested in real estate and 15% in commodities, respectively.

The paper is organized as follows. Section 3.2 explains the Minimax decision and Section 3.3 presents our data set. Section 3.4 provides our empirical results and Section 3.5 provides robustness checks for the results. Finally, Section 3.6 concludes.

3.2 Portfolio Selection Based On Minimax Decision Rule

Minimax optimal portfolios provide the best worst-case behavior. For a certain year with T trading days, let there be n risky assets with random daily returns $r_t^{(i)}$ $(i = 1, ..., n \text{ and } t = 1, ..., T)$. Further we assume that the investor holds a fraction w_i invested in each asset. The vector of weights $w = (w_1, ..., w_n)'$ should satisfy $w'\iota = 1$

(with $\iota = (1, 1, ..., 1)'$ being an $n-$dimensional vector of ones), i.e. all weights should add up to one. Since our representative investor is facing regulatory constraints, such that short-selling is not allowed for pension funds[3]. Hence, in the following, we prohibit short-selling (i.e. $w_i \geq 0$ for all i). Still, we relax this constraint later on, and we examine the performance of all trading strategies in the presence of possible short-selling.

Let $w'r_t = \sum_{i=1}^{n} w_i r_t^{(i)}$ be the portfolio payoff at day t, given the asset return vector $r_t = (r_t^{(1)}, \ldots, r_t^{(n)})'$ and a certain asset allocation w. To determine suitable weights for time t, a risk averse investor may solve the following Minimax optimization problem:

$$w^{\star} = \arg\max_{w} \min_{\tau \in \{t-1,\ldots,t-T\}} w'r_{\tau} \tag{3.1}$$

subject to $w'\iota = 1$ and $w_i \geq 0$. He considers the past T returns. In our application we set $T = 250$ to estimate the Minimax weights based on one year of daily returns. For any given asset allocation w, there is a worst daily outcome of $w'r_{\tau}$ with $\tau \in \{t-1, \ldots, t-T\}$. The Minimax portfolio is the asset allocation $w^{\star}$ which corresponds to the *max*imum return (or smallest loss) in the class of *min*imum outcomes.

3.2.1 Intuition

To make Minimax optimizations more intuitive, in the following we want to show how the Minimax procedure works with only two asset classes. If we look at the upper panel of Figure 3.1, we can see the possible portfolio return outcome at one specific day. All possible portfolio allocations are located on the straight line. Either we put all invested funds into stocks, or we invest all funds into bonds, or we spread the funds across both assets. At this particular day, stock performance was better than bond performance. If one considers this day only, one would be invested in stocks only. Now we look at another possible historical trading day (middle panel of Figure 3.1), and we see that at this particular trading day, stock performance was worse than bond performance. Now consider first the two extreme positions, being completely invested in stocks and being completely invested in bonds. If one is completely invested in stocks, the worst outcome occurs in the second day and corresponds to a gross return of about 0.85. One can improve this worst outcome by investing more into bonds. The worst outcome holding only bonds occurs in the first day with a gross return of 0.75. Again, one would be better off holding a more

[3] Literature on Mean-Variance optimization also suggests to impose short-sale constraints to reduce estimation error (e.g. see Frost and Savarino, 1986).

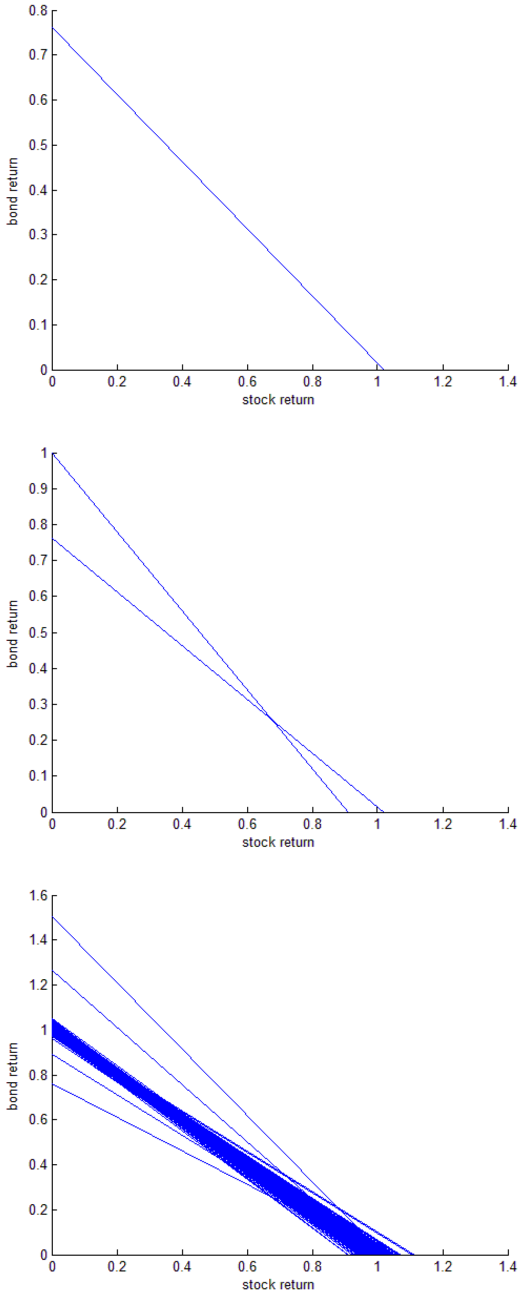


Figure 3.1: Illustration of Minimax for two Asset Classes

The figures illustrate the Minimax procedure for two asset classes only. The axis denote the change of the asset, B_t/B_{t-1} (y-axis) and S_t/S_{t-1} (x-axis), resp. The connecting line represents the outcome for any linear combination of both assets.

diversified portfolio. To sum up, in this two asset example Minimax optimization would take the diversified asset allocation that corresponds to the intersection of the two lines.

In the last panel of Figure 3.1, we now see all possible portfolio allocations for the last 250 trading days. For any given asset allocation the worst outcome corresponds to the point with the lines closest to the point of origin. Going again from the extreme portfolios towards more diversified portfolios (from both sides), we end up in the scenario that yields the highest portfolio return, given the worst case scenario. In our case the Minimax portfolio is the asset allocation which corresponds to the returns at the kink, i.e. at about the coordinates (0.65/0.25). We find that excluding one single day would not alter the optimal weights strongly. The Minimax leads to stable weights even though it depends on the single worst outcome. The optimization procedure is repeated each year to determine the weights for the following year.

3.2.2 Relationship to the Minimum Variance Portfolio

Consider an i.i.d. return time series of length T, $\{r_t\}_{t=1}^{T}$. What is the minimum over the T random variables? For i.i.d. normal distributed variables, Young (1998) considers an approximation based on the quantile, i.e. $\min_{t\in\{1,\ldots,T\}} r_t \approx \min_{t\in\{1,\ldots,T\}} E[r_t] \approx \mu + \Phi^{-1}(\frac{1}{T+1})\sigma$ with Φ^{-1} being the inverse of the standard normal cumulative distribution. Similar approximations can also be used for other distributions, such as the $t-$distribution. We assume that the minimum can be calculated by some general function f with respect to the mean μ, variance σ^2 and length of the time series T; i.e.

$$\min_{t\in\{1,\ldots,T\}} r_t \approx f(\mu, \sigma^2, T)$$

where returns r_t are i.i.d. distributed with first moments μ and σ^2. Function f can be a quantile approximation, an analytical solution for a certain distribution or some function based on numerical simulations.

Now consider a combination of n returns $w'r_t$ with $r_t = (r_t^{(1)}, \ldots, r_t^{(n)})$, mean $\mu = (\mu^{(1)}, \ldots, \mu^{(n)})$ and $n \times n$ covariance matrix Σ. The mean and covariance of portfolio w are given by $w'\mu$ and $w'\Sigma w$, respectively. If the investor applies the Minimum Variance portfolio, he assumes that the returns have a similar magnitude, i.e. $\mu^{(i)} = \mu_0$ for $i \in \{1, \ldots, n\}$. Under the given assumptions, the Minimax portfolio can be shown to be approximately equivalent to the Minimum Variance portfolio.

$$\begin{aligned}
w^\star &= \arg \max_{w, w'\iota=1} \min_{t\in\{1,\ldots,T\}} w'r_t \\
&\approx \arg \max_{w, w'\iota=1} f(w'\mu, w'\Sigma w, T) \\
&= \arg \max_{w, w'\iota=1} f(\mu_0, w'\Sigma w, T) \\
&= \arg \min_{w, w'\iota=1} w'\Sigma w
\end{aligned}$$

We have seen that for similar means and a "nice" distribution of the returns the Minimax portfolio is similar to the Minimum Variance portfolio.

But why should an investor apply the Minimax portfolio instead of Minimum Variance portfolio? Mainly there are three advantages. First, opposed to the Minimum Variance portfolio, the Minimax portfolio uses an asymmetric risk measure and does not penalize positive outliers. Second, the Minimax asset allocation model adjusts quickly to structural breaks. Third, it implicitly considers the mean and not only the variance. We present the behavior by a structural break to (i) the mean and to (ii) the variance. The impact on the weights is shown in Figure 3.2. We consider an example of two assets with daily returns $r_t^{(1)}, r_t^{(2)}$. The uncorrelated returns have equal means[4] $\mu^{(1)} = \mu^{(2)}$ and variances $\sigma^{(1)} = \sigma^{(2)} = \sqrt{10^{-5}}$. The weights are estimated based on an estimation window of one year (250 days). As the returns are equal, the optimal weights are given by $w^\star = \left(\frac{1}{2}, \frac{1}{2}\right)'$. In the period from year 1 to year 2 the Minimax portfolio as well as the Minimum Variance portfolio are close to the theoretical optimal weights $w^\star$. The top plots of Figure 3.2 shows the reaction of a large structural break to the mean for normal (left) and t_3-distributed returns (right). At year 2, the mean return of asset 1 drops by 10^{-3} on a new level. We observe that the Minimax weight in asset 1 is immediately decreasing. After less than a month already more than half of the weight adjustment has been done. As the Minimax portfolio considers tail events, the break in the mean has a smaller impact on the weights, if the returns have fatter tails. The weight adjustment for the $t-$distributed returns (right plot) is smaller than for the normal distributed returns (left plot). The Minimum Variance portfolio decreases slowly its share in asset 1. Note that the decrease does not stem from the fact that the variance has changed (it has not) but from the bias in the mean estimate. At year 2.5, half of the returns of the estimation window have the initial mean (before year 2) while the other half of the returns have the mean after the structural break (after year 2). The sample variance is high and the Minimum Variance portfolio invests less in asset 1. As the estimation window is one year, after year 3 there is no more bias and the Minimum Variance allocation is again the initial allocation: $w = \left(\frac{1}{2}, \frac{1}{2}\right)'$. The Minimax optimization continues to hold a smaller share in asset 1. After year 2, asset 1 has a smaller mean than asset 2, while both have the same variance. The weights of the Minimax are hence more reasonable than the Minimum Variance weights.

The lower plots of Figure 3.2 show the behavior of the portfolio strategies if a structural break to the variance occurs. Before year 2, the returns are distributed as before with same mean and variance. At year 2, a structural break to the variance of asset 2 occurs, i.e. $\sigma_t^{(2)} = \sqrt{2 \cdot 10^{-5}}$ for day t after year 2. For normal returns, we find that the Minimax portfolio and the Minimum Variance portfolio behave similarly. Both portfolios adjust over a one year period to new weights $w = \left(\frac{2}{3}, \frac{1}{3}\right)'$. However, the Minimax portfolio adjusts at a slightly quicker rate, with about one month advance to the Minimum Variance portfolio. For $t-$distributed returns, the reaction of the Minimum Variance portfolio remains the same. The Minimax port-

[4] As the means are equal, the magnitude does not influence the weights of the considered portfolio strategies.

folio adjusts less than before to $w \approx (0.6, 0.4)'$. This feature seems not to be bad either. For (possibly wrongly) estimated extreme variances, the Minimax portfolio remains stronger diversified.
We find that the Minimax reacts quicker to structural breaks and avoids misspecifications as it also considers the mean. The cost of the Minimax portfolio (as well as all portfolios considering tail risk) are more unstable weights. The behavior is reflected by greater (bootstrapped) confidence bands of the weights. However, the confidence bands of the Minimum Variance portfolio are strongly increasing as the distribution of returns becomes more fat-tailed. The confidence bands increase less for the Minimax portfolio. That means in case of structural breaks and fat-tailed returns, the Minimum Variance portfolio may be as instable as the Minimax portfolio. Then the Minimax portfolio is preferable due to its better characteristics.

3.2.3 Benchmark Models for Minimax Trading Strategy

We test several benchmark models to compete against the proposed Minimax trading strategy. The benchmark models are chosen according to their relevance in the literature. We additionally show results for the "fixed weights" strategy, in order to mimic a representative investor's trading strategy. This fixed weights strategy is based on average holdings of US pension funds and invests 60% into stocks, 20% into bonds, 10% into real estate and 10% into commodities, respectively.
A strategy that is closely related to the fixed weights strategy is the equal weights strategy. Here we apply the naive $\frac{1}{n}$ rule of thumb to allocate funds to different asset classes. Since $n = 4$ in our case, we invest 25% of total funds into each asset class every year. One advantage of both strategies is obvious: Transaction costs are negligible. And as DeMiguel et al. (2009b) state, the naive $\frac{1}{n}$-rule is still hard to beat for any optimization strategy, since estimation risk does not exist.
Now we introduce the Mean-Variance related trading strategies that we are going to test our model against in the following section. First of all, we introduce the Minimum Variance portfolio as the competitor that is most likely to beat the Minimax trading strategy for the following reasons. A body of literature is written on problems in Mean-Variance optimization due to estimation of expected returns. Small deviations of expected returns from realized returns result in high fluctuating weights (Jorion, 1991). Merton (1980) proposes to ignore expected returns and use the Minimum Variance portfolio instead. It minimizes the risk of a portfolio and therefore is a suitable strategy for any highly risk-averse investor. The main differ-

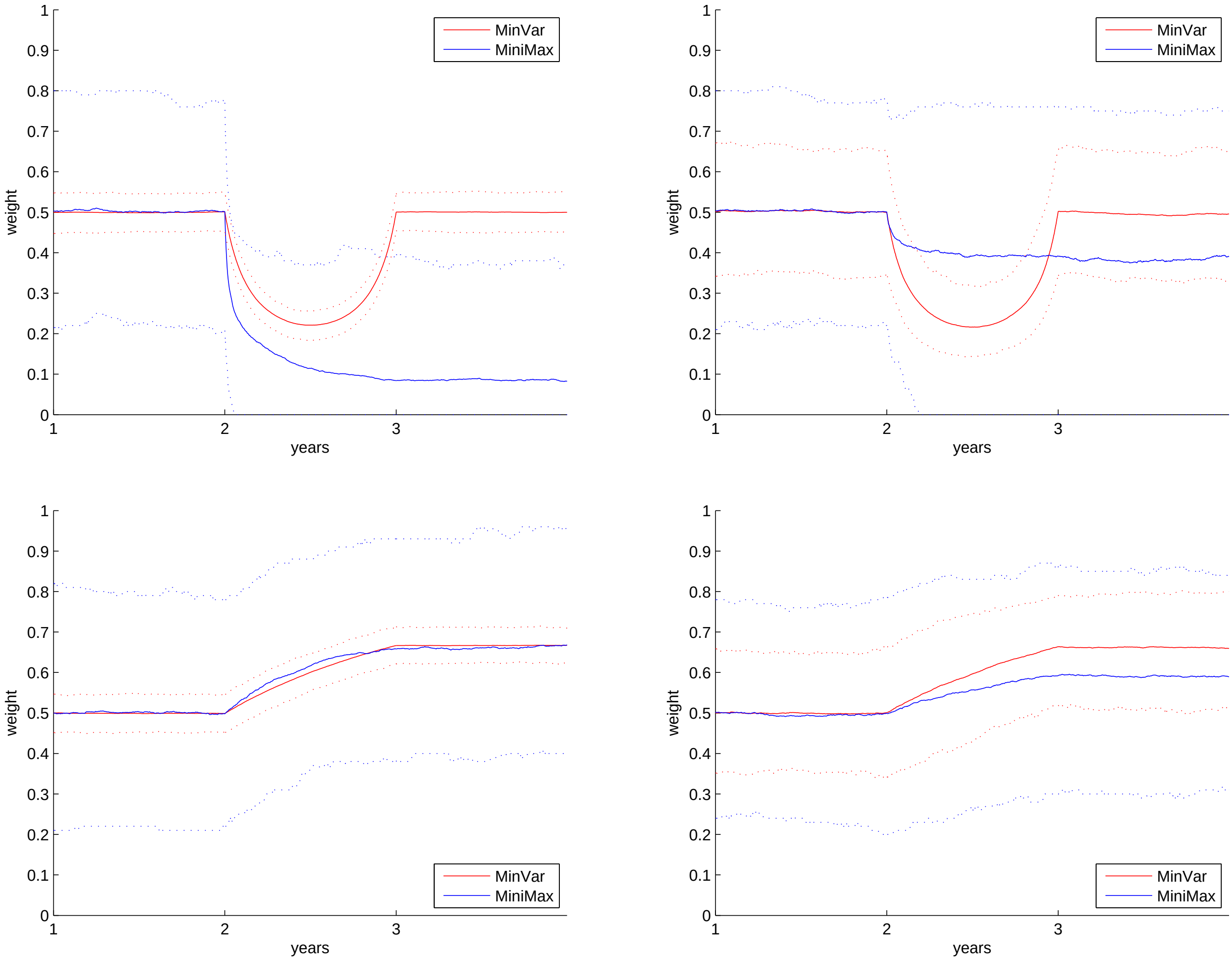


Figure 3.2: Structural Break to Moments: Minimax vs. MinVar

The figure shows the weights of the Minimax portfolio (blue line) and the Minimum Variance portfolio (MinVar, red line) under a structural break to the mean (top plots) and the variance (lower plots) at $t = 2$. The returns are either normal (left plots) or t_3 distributed (right plots). The $y-$axis represents the weight for asset 1 (the weight for asset 2 is hence given by $1 - y$). The $x-$axis denotes the time (in years). The returns of both assets $i \in \{1, 2\}$ are $N(\mu^{(i)}, \sigma^{(i)})$ distributed with $\mu^{(1)} = \mu^{(2)}$ and $\sigma^{(1)} = \sigma^{(2)} = \sqrt{10^{-5}}$. At year 2 there is a lasting shift of the mean $\mu^{(1)} = \mu^{(2)} - 10^{-3}$ (top plot) and the variance $\sigma^{(2)} = \sqrt{2 \cdot 10^{-5}}$ (lower plot), respectively. The bootstrapped 90% confidence bands are given by the dotted lines.

ence to Minimax is the risk measure, which is symmetric (volatility). In particular, the Minimum Variance optimization is given as

$$\underset{w}{min}\, w'\Sigma w \tag{3.2}$$

subject to $w'\iota = 1$ and $w_i \geq 0$, with Σ being the 4×4 covariance matrix of our four asset classes.

The last strategy that we want to test our Minimax strategy against is the Mean-Variance strategy. It is based on Markowitz (1952) and relates risk and return as given in the following optimization:

$$\underset{w}{max}\, w'\mu - \frac{1}{2}\lambda w'\Sigma w \tag{3.3}$$

subject to $w'\iota = 1$ and $w_i \geq 0$. We calculate Mean-Variance optimal portfolios based on a risk-aversion parameter of three, i.e. $\lambda = 3$ (as commonly chosen, see e.g. Kan and Zhou, 2007).

3.3 Data

To form portfolios, we use the same indexes as Hodder et al. (2012). It consists of four indexes on major asset classes that big US pension funds commonly and mainly invest in: stocks, bonds, real estate and commodities. We notice that a major share of pension funds actually invests in more asset classes (e.g. in hedge funds). We do only consider stocks, bonds, real estate and commodities for two reasons: First, the share invested in alternative asset classes is relatively small with about 5%. Second, in our Minimax optimization, we are limited to a certain number of asset classes due to computational issues, and four asset classes turn out to be still feasible.

Our database is Thomson Datastream, and we download daily prices of performance indexes on the four asset classes. For stocks, we consider the total return (dividends included) on the S&P 500 index. Bond performance is measured by the Barclays Aggregate Bond index, also in terms of a total return index. For real estate, we use the Datastream US real estate index, which is an appropriate proxy for the US real estate market. We note that pension funds also may directly own residential and commercial real estate. However, we lack daily prices and therefore rather consider the performance of a real estate investment trust, where daily valuation is available. Commodity market performance is represented by the common used S&P GSCI index, which is widely diversified across commodities. Finally, we take the yields on a 90-day Treasury Bill as a proxy for the risk-free rate to calculate the Sharpe ratio. Our indexes cover 21 years from January 1990 to December 2010, resulting in 5480 daily observations. We provide the performance of all asset classes in Figure 3.3 and descriptive statistics for all asset classes in Table 3.1.

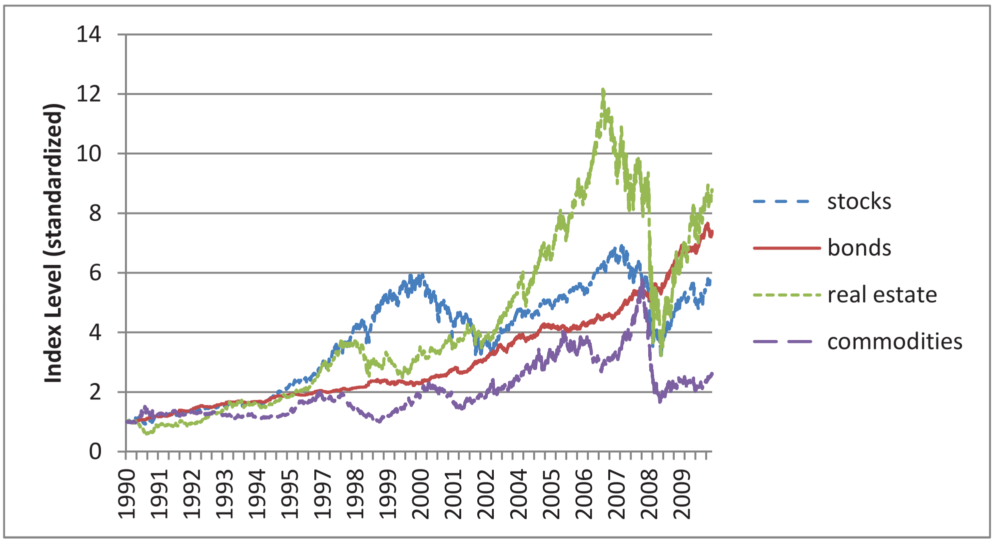


Figure 3.3: Performance of Asset Classes

The figure shows the time series of one dollar invested into each asset class in January 1990. Sample period is January 1990 to December 2010.

Table 3.1: Descriptive Statistics

PERFORMANCE	Geometric Return (%)	Volatility (%)	Sharpe Ratio
Stocks	0.033	1.15	0.022
Bonds	0.037	0.43	0.055
Real Estate	0.040	1.66	0.024
Commodities	0.018	1.37	0.009

DISTRIBUTION	Geom. Return (%)	Min (%)	Max (%)	Skewness	Kurtosis
Stocks	0.033	-9.0	11.6	-0.01	9.46
Bonds	0.037	-4.0	5.1	-0.66	8.29
Real Estate	0.040	-18.3	18.6	+0.48	23.53
Commodities	0.018	-16.8	7.9	-0.41	7.48

CORRELATION	Stocks	Bonds	Real Estate
Bonds	-0.07		
Real Estate	0.65	-0.01	
Commodities	0.11	0.03	0.09

All numbers are based on daily returns. Returns (%) are calculated geometrically. The Sharpe Ratio is calculated by subtracting the daily risk-free rate from daily returns, and then dividing by their daily standard deviation. Min (%)/Max (%) is the minimum return/maximum return of all realizations in the sample period, denoted in percentage. Correlations are calculated unconditionally on the returns. All measures are calculated over the whole period of 1990 to 2010.

We note that in the considered time period, real estate has been the asset class with the highest return, but also with the highest volatility. Especially during the financial crisis in 2008 we notice a heavy decline of US house prices. Also stock and commodity prices decreased during that time, in contrast to US bonds, which incorporate the lowest volatility of all four asset classes. Their volatility is only as high as 0.43% for daily prices, which is by far less than half of the volatility of stocks, real estate and commodities. Another risk measure, minimum return, indicates that bonds are less risky than stocks, real estate and commodities, losing only 4.0% in the worst trading day over the whole period. Bonds are also attractive, when we relate

risk and return, looking at the Sharpe Ratio. Again, bonds outperform other asset classes. We can thus expect that bonds will play a major role in all risk-reducing optimizations.
When we look at correlations between asset classes, we find that almost all asset classes exhibit low correlations. The only exception is a relatively high correlation between stocks and real estate. We conclude that this is because our real estate investment trust (REIT) is exchange-traded. We can thus further expect that a risk-reducing optimization will diversify funds across asset classes, since diversification potential is high due to low asset correlations.

3.4 Empirical Study

3.4.1 Minimax Performance

We now show results for our empirical analysis. Each year in January, we run the Minimax and Mean-Variance optimizations based on the last 250 trading days. We obtain optimal weights for each optimization strategy, and we hold a portfolio based on these optimal weights for one subsequent year[5]. After this year, we evaluate the (out-of-sample) performance and, we rerun the optimization based on the daily return realizations of the particular year, and so on. Furthermore, we show results for the equal weights and fixed weights strategies, which merely reallocate funds to their predetermined shares. Table 3.2 presents results of performance and risk for all trading strategies.
Looking at the out-of-sample performance of our portfolios, we can see that the Mean-Variance portfolio (with $\lambda = 3$) yields the highest return, but also the highest volatility. Neglecting the Mean-Variance portfolio, with an average daily return of 0.05% and a volatility of 0.54%, the Minimax strategy has the highest Sharpe Ratio. We can also see that the Minimum Variance portfolio has the lowest (out-of-sample) volatility, which is in line with the objective of Minimum Variance optimization, i.e. to minimize (in-sample) volatility. Interestingly, by diversifying funds widely across all asset classes, the naive strategies beat the Mean-Variance portfolio in terms of volatility; still the Minimax and the Minimum Variance strategies outperform naive diversification strategies not only in terms of risk, but even yield a higher return. All portfolios (but the Mean-Variance portfolio) have lower Sharpe Ratios with re-

[5]In a robustness check, we also provide results for a rolling window approach, rebalancing our portfolio daily (in contrast to yearly).

Table 3.2: Performance and Risk Results

PERFORMANCE	Return (%)	Vola (%)	Sharpe Ratio	Excess Return	TO	CEV3
Minimax	0.052	0.54	0.071	over Minimax (%)	9.2	4.8
Minimum Variance	0.039	0.40	0.063	-0.013	9.9	3.7***
Equal Weights	0.040	0.78	0.032***	-0.012	0.0	3.0**
Fixed Weights	0.040	0.77	0.034***	-0.012	0.0	3.1**
Mean-Variance	0.105	0.97	0.094***	+0.050	26.4	9.1***

RISK	Volatility (%)	VaR(0.95) (%)	VaR(0.99) (%)	Max. Loss (%)
Minimax	0.54	-0.57	-0.99	-2.06
Minimum Variance	0.40	-0.54	-1.08	-2.63
Equal Weights	0.78	-0.95	-2.29	-8.71
Fixed Weights	0.77	-0.99	-2.07	-7.31
Mean-Variance	0.97	-1.71	-3.59	-9.25

All numbers are based on daily returns. Return (%) is calculated geometrically. The Sharpe Ratio is calculated by subtracting the daily risk-free rate from daily returns, and then dividing by their daily standard deviation. Turnover (TO) is measured by the sum of absolute deviations in weights before and after rebalancing. We rebalance portfolios every year, using past 250 trading days for our optimizations. CEV3 is the Certainty Equivalent for a Mean-Variance investor with risk aversion parameter $\lambda = 3$ scaled by 10^2. VaR stands for Value-at-Risk, which indicates the minimum return (or equivalently, the maximum loss) for a certain significance level (95% or 99%). Based on the test of Ledoit and Wolf (2008) significant Sharpe Ratio and CEV3 differences of the alternative portfolio model and the Minimax at 1%, 5%, 10% level are denoted by $\star\star\star$, $\star\star$ and $\star$, respectively. Short-selling is not allowed for all optimizations. All measures are calculated over the whole period of 1990 to 2010.

spect to Minimax, and the difference of Sharpe Ratios with respect to Minimax is significant for all portfolios, except for Minimum Variance. All results are obtained by imposing short-selling constraints, which implicitly enhance the performance of Mean-Variance related competitors; see, for example, Jagannathan and Ma (2003). However, we have to bear in mind that this gain comes at the cost of rebalancing the portfolio from year to year. Active rebalancing leads to higher transaction costs, which may compensate the benefits of optimization and which works in favor of easy rule-of-thumb strategies with low portfolio turnover. The strategies with less turnover are by construction the equal weights and fixed weights strategies, which cause almost no transaction costs. We proxy turnover by the sum of total deviation in weights from one period to another:

$$TO_t = \sum_{i=1}^{4} \mid w_{i,t} - w_{i,t-1} \mid \tag{3.4}$$

Yet, we neglect here that asset prices change over time. In reality even naive strategies have slightly positive transaction costs, since portfolios have to be reset to their predetermined fixed allocation from year to year. Mean-Variance portfolios have the highest turnover, which is more than twice as high as the turnover of Minimum Variance and Minimax strategies. This is not surprising, though. The portfolio literature shows that pure Mean-Variance related portfolio optimizations exhibit strongly fluctuating weights, which result in high transaction costs. Therefore, several approaches have been proposed to handle the problem of estimation error, which is mainly responsible for highly fluctuating weights. For instance, Ledoit and Wolf (2003, 2004a,b) suggest to use a weighted average of the sample covariance and a constant variance estimator. By doing so, extreme weights can be prevented and thus, transaction costs can be substantially lowered. For practical implementation, transaction costs play a major role. Remember that our representative agent is a big US pension fund investing billions of dollars for its clients. A small reduction in costs is equivalent to saving a huge amount of money.

Now it is interesting to compare turnover of our Minimax approach to Minimum Variance optimization. We can see that turnover is comparable, but still, turnover is slightly less for the proposed Minimax strategy.

Last but not least, comparing the Certainty Equivalent of all strategies, we find that the Mean-Variance approach yields the highest CEV3 value out-of-sample, which is not surprising, too, since the Certainty Equivalent is a concept originating in the Mean-Variance theory of Markowitz (1952). Again, CEV3 differences of all portfolios with respect to Minimax are significant. To sum up the performance of our four portfolios, we find that Minimax is always best or second best (but for turnover), and thus, Minimax is very consistent across all performance measures used. Now we want to stress one component of performance, which is highly relevant for a big US pension fund portfolio, namely its risk. Risk can be measured differently, and here we use a selection of risk measures that is often used in scientific work as well as in practical work. We use volatility as a common symmetric risk measure in finance. Additionally, we use the Value-at-Risk (VaR) measure as the most frequently used asymmetric risk measure. Finally, we use the maximum loss as an asymmetric risk measure that fits best our risk aversion assumption, namely that our representative investor is highly risk-averse, and as a result, he fears high losses.

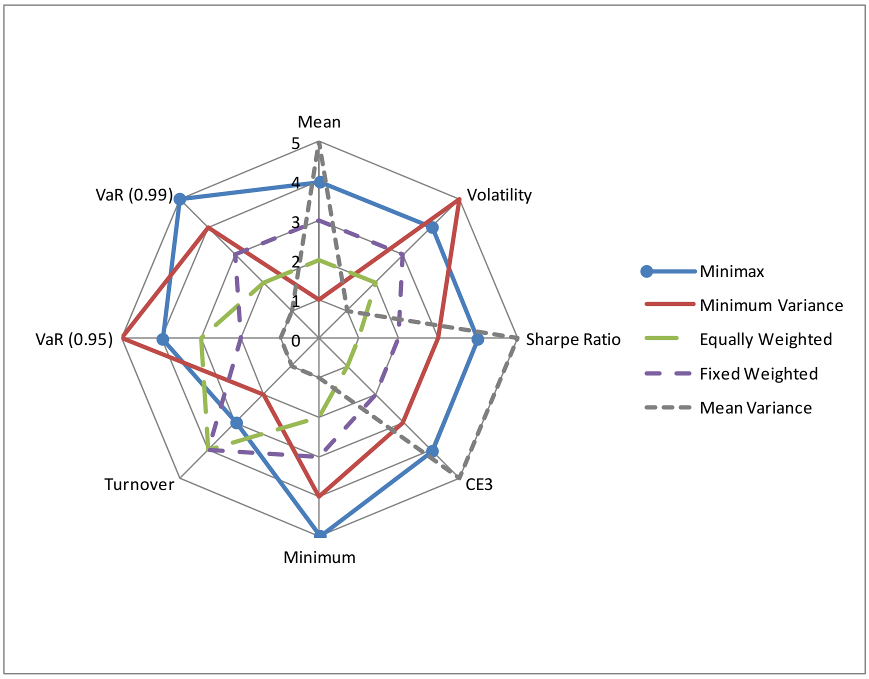


Figure 3.4: Graphical Illustation of Performance and Risk Results

All performance and risk results are based on the numbers and variables in Table 3.2. We rank all strategies, while 1 is given to the strategy that performs the poorest for a particular risk measure, and 5 is given to the best strategy, respectively. All measures are calculated over the whole period of 1990 to 2010.

Consider the risk analysis based on the Value-at-Risk measure. The Value-at-Risk indicates the maximum loss that is possible for a given significance level, here we choose the level according to 95% and 99%. We find that on the 95% significance level, the Minimum Variance portfolio is the one with the lowest maximum loss. Strategies of representative pension funds are exposed to risk, which is almost twice as high as Minimum Variance and Minimax optimizations, based on the VaR(0.95) risk measure. The results are basically the same when we look at the Value-at-Risk at the 99% significance level. Again, naive diversification strategies load twice as much risk than risk-minimizing strategies. The Mean-Variance portfolio is not competitive based on any risk measure. Interestingly, the Minimum Variance portfolios outperform Minimax portfolios at the 95% VaR significance level, but it appears to be the other way round at the 99% significance level. The same result is obtained for maximum loss, where the Minimax portfolio clearly dominates all other strategies. We can see from the results that for a highly risk-averse investor, Minimum Variance is the hardest competitor for our Minimax model.

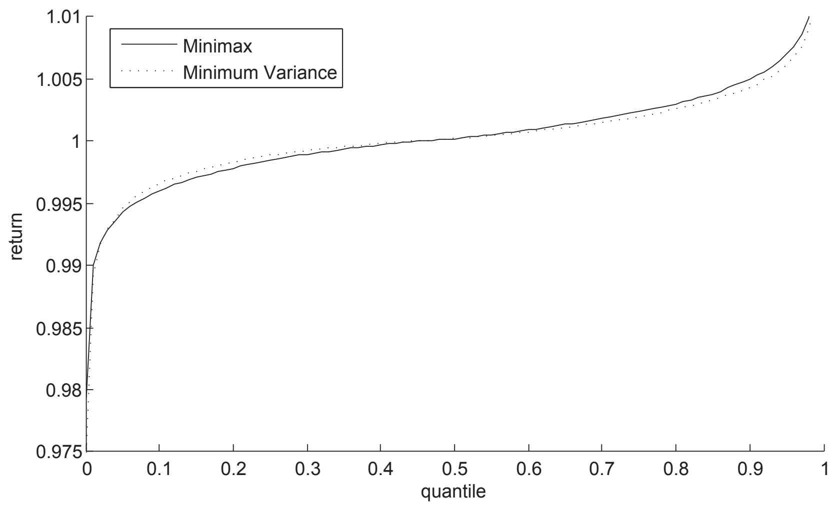


Figure 3.5: Empirical Quantiles of Minimax and Minimum Variance

The figure shows the portfolio return distributions of Minimax and Minimum Variance strategies. The quantiles are given on the $x-$axis. The returns are given on the $y-$axis, with 1.00 representing a zero portfolio return. The portfolio returns are calculated over the whole period of 1990 to 2010.

The last result indicates that Minimax portfolios outperform its closest competitor in the very left tail of return distributions, i.e. for a highly risk-averse agent like ours. We want to examine this issue by looking at the return distributions of Minimax and Minimum Variance portfolios. Figure 3.5 shows both distributions. Looking at very low portfolio returns (i.e. at the very left hand of Figure 3.5), we notice that Minimax dominates Minimum Variance. In other words, for the portfolio risk that matters most for a highly risk-averse agent, Minimax outperforms the Minimum Variance strategy. For low, but not extreme returns we can see that it is the other way round, while turning to positive returns, again Minimax dominates Minimum Variance. This is because Minimum Variance minimizes a symmetric risk measure, volatility, and therefore chooses a portfolio with returns that do not deviate too far from their mean. This property is, however, not desireable. Imagine a portfolio A which is more volatile than portfolio B. Assume that the returns of A are constantly higher than the returns of B. This implies that the realized loss of B is always larger than the loss of portfolio A. Still, the Minimum Variance strategy would always prefer portfolio B to A, although logically, a risk-averse investor with any reasonable utility function should prefer portfolio A to B (and this is what Minimax would do). The outperformance of Minimax over Minimum Variance in terms of the commonly used Sharpe Ratio can thus be summarized as a dominance of Minimax in the right-

Table 3.3: Portfolio Characteristics

WEIGHTS	Min. All	Max. All	Average (Vola) Stocks	Average (Vola) Bonds	Average (Vola) Real Estate	Average (Vola) Commodities
Minimax	0.0	0.9	0.11 (0.11)	0.54 (0.24)	0.15 (0.18)	0.15 (0.11)
Minimum Var	0.0	0.9	0.15 (0.10)	0.61 (0.26)	0.09 (0.14)	0.10 (0.10)
Equal Weights	0.25	0.25	0.25 (0.00)	0.25 (0.00)	0.25 (0.00)	0.25 (0.00)
Fixed Weights	0.1	0.6	0.60 (0.00)	0.20 (0.00)	0.10 (0.00)	0.10 (0.00)
Mean-Variance	0.0	1.0	0.10 (0.30)	0.10 (0.30)	0.43 (0.51)	0.33 (0.48)

CORRELATION	Minimax	Minimum Variance	Equal Weights	Fixed Weights
Minimum Var	0.73			
Equal Weights	0.54	0.52		
Fixed Weights	0.50	0.52	0.52	
Mean-Variance	0.25	0.24	0.78	0.24

All figures in the upper table are based on asset weights resulting from portfolio optimizations. “Minimum All” and “Maximum All” denote minimum and maximum weights for any asset class, whereas “Average” denotes the average weight in one specific asset class. Correlation measures in the table below are again based on daily portfolio returns. All measures are calculated over the whole period of 1990 to 2010.

hand part (positive returns) of return distributions as well as in the very left part (extreme losses).

3.4.2 Analysis of Portfolio Weights

Besides attractive performance characteristics, one should also look at other portfolio characteristics. A big pension fund aims at maintaining rather constant portfolio weights than at rebalancing frequently due to higher transaction costs. We already considered turnover as a proxy for transaction costs, and we found that the naive strategies cause the least transaction costs, Mean-Variance optimization leads to the highest transaction costs, and turnover of Minimum Variance and Minimax strategies is quite comparable. This fact leads to the question, to what extent both best performing strategies are similar. When we consider correlations of portfolio returns, we can see that Minimax and Minimum Variance portfolio returns are highly positively correlated with a correlation coefficient of 0.73. Only one other pair of portfolio strategies is as highly correlated. The high correlation indicates that the risk minimizing mechanisms must be similar. The finding is not surprising. In section 3.2.2 we have seen that under certain conditions, the Minimum Variance Portfolio and the Minimax Portfolio are based on similar optimization procedures. Both Minimax and Minimum Variance hold the largest fraction of funds in bonds. Before we saw that bonds have quite nice performance characteristics; they have the highest Sharpe Ratio and the lowest volatility. Thus, it is not surprising that, on average, a 54% share of Minimax portfolios and a 61% share of Minimum Variance portfolios is invested into this asset class. In contrast, Minimax invests a little more into real estate (15%) and commodities (15%) than the Minimum Variance strategy (9% and 10%, respectively). The Mean-Variance strategy is on average heavily invested into real estate and commodities, while their share in stocks and bonds is only 20%. This, however, is a result of fluctuating weights in the Mean-Variance optimization, as we can see in Figure 3.6. Mean-Variance portfolios always consist of only one asset class, which is held for one period. In contrast, Minimax and Minimum Variance strategies seem to be quite stable over time. Figure 3.6 plots the time series of portfolio weights.

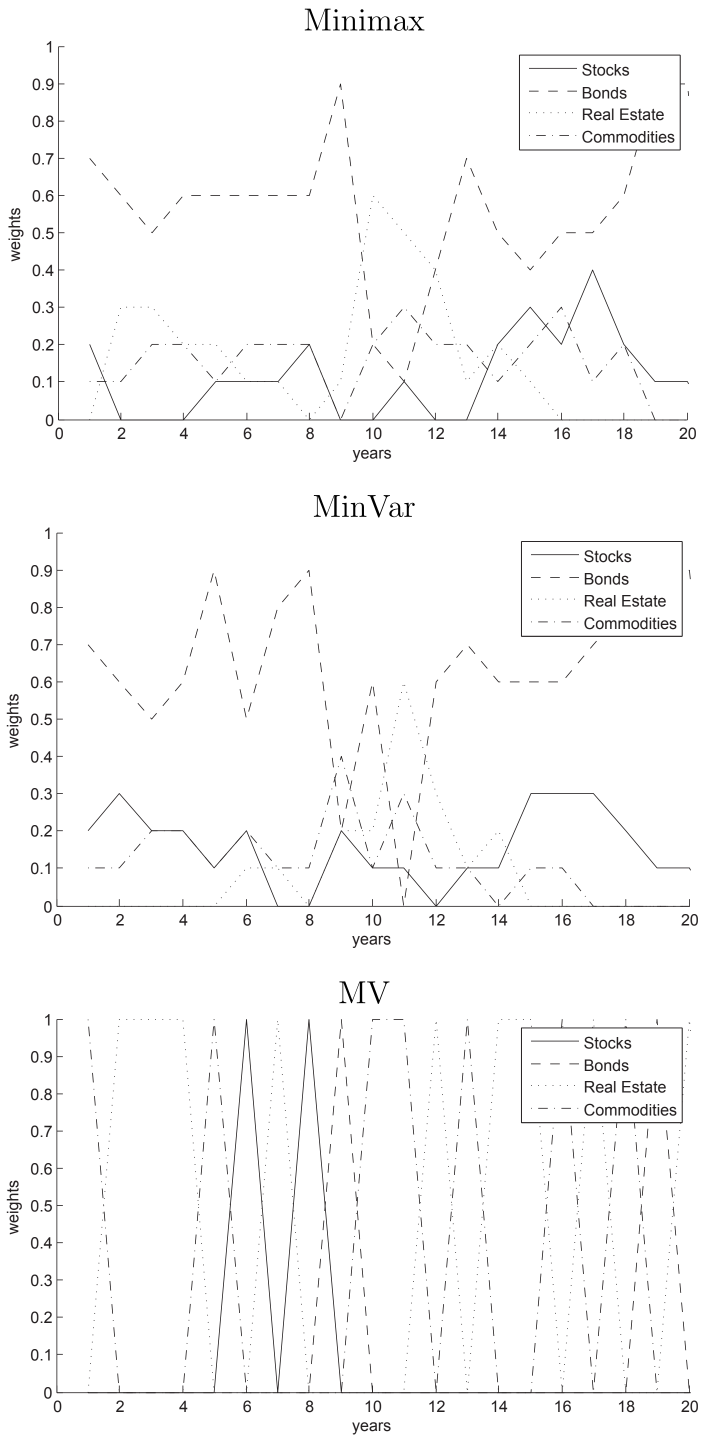


Figure 3.6: Time Series of Portfolio Weights

This figure shows portfolio weights for each asset class, resulting from Minimax, Minimum Variance and Mean-Variance optimizations. Rebalancing takes place every year. The number 1 on the abscissa corresponds to year 1991, since we need one year of historical returns (January 1990 to December 1990) for the optimizations.

3.5 Robustness Checks

In this section, we check our results for robustness. To do so, we slightly change Minimax as well as other optimization procedures, and we see that results basically stay the same.
First, we rerun optimizations and fixed weight strategies for yearly returns, in contrast to daily returns. We provide results for the main performance measures in 3.A.1. Table 3.A.4 shows that when we look at annual returns, the Minimum Variance portfolio dominates Minimax both in terms of risk and return. This is due to the return distribution. On the one hand, yearly returns are closer to the normality assumptions than daily returns, and thus, Minimax optimization loses its advantage in the very left part of the return distribution, because high portfolio losses are rare. On the other hand, the closeness to normality of portfolio return works in favor of optimizations using volatility as a symmetric risk measure. However, we see that Minimax is able to beat naive asset allocation strategies and therefore, is preferable from the point of view of big US pension funds. Again, we want to highlight the advantage of Minimax for daily periods of interest, for instance for institutional investors that are due to daily risk reporting.
As a second robustness check, we want to consider a different rebalancing period. Up to now, we rebalance portfolios once a year, how it is commonly done in the literature, and our estimation window consists of 250 past observations. Now we want to consider a rolling windows approach: We rebalance every day, and again, the estimation window consists of 250 past observations. Results are provided in 3.A.2. Table 3.A.5 shows that Minimax has the highest Sharpe Ratio. Again, based on the Sharpe Ratio, we find that the Minimax portfolio outperforms all considered benchmarks. For Minimax portfolio returns, volatility is slightly higher than for Minimum Variance portfolios, still returns are again higher. All other results stay qualitatively the same as in the base scenario with yearly rebalancing.
In a third robustness check, we allow for short-selling, i.e. we do not require the asset weights to be positive. This obviously works against Mean-Variance related optimizations that suffer from estimation error. We can see results for optimizations in 3.A.3. Table 3.A.6 shows performance and risk characteristics for our standard case, daily returns and yearly rebalancing. For Minimum Variance portfolios, allowance for short-selling does not change. For Mean-Variance portfolios, however, we can see that the portfolio risk increases substantially, as well as turnover, which is a result of large long and short positions in the portfolio. Such a portfolio is not

preferable for our typical investor, since it holds substantial risk, and turnover is around 25 times higher than for Minimum Variance or Minimax portfolios. However, the Mean-Variance portfolio without short-selling restriction yields a Sharpe Ratio which is higher than that of its competitors. Again, the Sharpe Ratio for our Minimax portfolio is higher than that of Minimum Variance. Table 3.A.7 provides robustness checks for daily returns and daily rebalancing without short-selling restrictions. As expected, turnover increases in addition for all portfolios and yields 3818.9 for the Mean-Variance portfolio, as opposed to around 40.4 and 46.5 for Minimax and Minimum Variance portfolios, respectively. Again, the Sharpe Ratio of Minimax exceeds the Sharpe Ratio of the Mean-Variance related strategies.
The last robustness check is with respect to crises periods. In 3.A.4, we analyze the performance of the Minimax portfolio and its competitors during crises periods. We consider the following crises: (1) the Internet bubble during 2000/2001, (2) the Terrorist attack in 9/11/2001, (3) the Banking crises during 2007/2008 and subsequent (4) Economic recession during 2008-2010. The Minimax portfolio suffers of the smallest maximum loss in all crises but one. In all considered crises it has the largest or second largest CE3 performance. In all cases it achieves Sharpe ratio which is above average - sometimes best. We conclude that in particular for crises periods, the Minimax portfolio performs well.

3.6 Conclusion

In this paper, we propose a trading strategy called Minimax, which is based on pessimistic decision making and which suits a highly risk-averse investor. In particular, Minimax is an appropriate asset allocation optimization for big pension funds or other institutional investors that are due to daily risk reporting, either because of regulatory requirements or because of mark-to-market accounting. Maximizing the worst case payoff of a portfolio, Minimax strategies are practically easy to implement and constitute a proper alternative to common risk-minimizing optimizations such as Minimum Variance.
We use US data on indexes of stocks, bonds, real estate and commodities from January 1990 to December 2010 in order to calculate daily portfolio returns. We compare the proposed allocation strategy with alternative asset allocation strategies. Therefore, we calculate a Minimum Variance portfolio, which minimizes volatility within one year of historical daily returns, a Mean-Variance portfolio, an equal

weights strategy, and a typical US pension fund portfolio, which imitates an asset allocation that a representative investor could possibly run.
Our main result is that the proposed Minimax strategy outperforms all competitors, in terms of different risk and performance measures. We find the Minimum Variance portfolio to be the hardest competitor for Minimax, assuming a highly risk-averse agent, and portfolio characteristics of both strategies are comparable. We show that the particular advantage of the Minimax strategy is the avoidance of very large losses. Optimizations based on volatility as a symmetric risk measure such as Minimum Variance strategies fail to provide optimal portfolios with attractive performance characteristics, because they minimize not only negative, but also positive returns. Minimax, however, only cares about huge portfolio losses, and provides preferable performance characteristics by allowing positive portfolio returns. Naive portfolio allocation rules are not competitive to Minimax in terms of performance and risk. Still, this does not mean that naive portfolio allocation strategies are in general not appropriate for any investor. Studies show that many optimization strategies fail to beat simple rules of thumb, and by diversifying funds across different asset classes, one can reduce risk efficiently without imposing strong ex ante restrictions. Another advantage of fixed weight strategies, besides their simplicity and applicability, is low turnover. This feature makes them particularly interesting for long-term investors that face high transaction costs. Mean-Variance optimization is not competitive to all above mentioned strategies, due to high estimation error.
Considering portfolio characteristics, we find admirable features for portfolio weights that result from all strategies but from Mean-Variance optimization. Portfolio weights are relatively stable over time for Minimax and Minimum Variance, resulting in comparable turnover and transaction costs. Transaction costs are particularly high for the Mean-Variance portfolio, as one would expect given high estimation errors and extreme portfolio weights. Minimax and Minimum Variance portfolios invest on average about 50-60% of their funds into bonds, while the remaining capital is spread across the remaining asset classes. By doing so, the resulting portfolios are satisfyingly diversified.
In a last analysis, we check our results for robustness. For yearly returns (instead of daily returns), we lose dominance over the Minimum Variance strategy. This is particularly due to the fact that yearly returns are “closer to normality” than daily returns. In a scenario with normally distributed returns, portfolios based on Markowitz (1952) are shown to be optimal. We also check whether the chosen re-

balancing period of one year has particular influence on our results. We show that using daily rebalancing, we obtain even better results as when using yearly rebalancing. Lastly, we allow for short-selling, which was constrained in all optimizations before. As expected, Mean-Variance weights fluctuate even more, resulting in high turnover and transaction costs. The dominance of Minimax over Minimum Variance portfolios remains.
All results suggest that Minimax strategies provide an attractive alternative asset allocation optimization for a highly risk-averse investor that is concerned with daily risk management. Since Minimax prevents portfolios from realizing high extreme losses, institutional investors that are due to daily risk management can lower their daily portfolio risk. Additionally, Minimax strategies are easily implementable due to its simple algorithm and because exchange traded funds provide easy access to all considered asset classes.

Appendix 3.A Robustness

3.A.1 Robustness Check: Yearly Returns

Table 3.A.4: Robustness Check: Yearly Returns

PERFORMANCE	Return (%)	Vola (%)	Sharpe Ratio	Excess Return	TO	CEV3
Minimax	11.2	10.3	0.739	over Minimax (%)	8.2	9.6
Minimum Variance	11.9	9.3	0.9⋆⋆⋆	0.7	8.9	10.6⋆
Equal Weights	11.7	12.4	0.653	0.5	0	9.4
Fixed Weights	11.4	12.3	0.631	0.2	0	9.1
Mean-Variance	11	23.3	0.318	-0.2	30	2.9

RISK	Volatility (%)	VaR(0.95) (%)	VaR(0.99) (%)	Max. Loss (%)
Minimax	10.3	-8.10	-10.8	-10.8
Minimum Variance	9.3	-2.4	-4.3	-10.0
Equal Weights	12.4	-17.6	-28.3	-28.3
Fixed Weights	12.3	-14.9	-23.6	-36.9
Mean-Variance	23.3	-37.6	-45.0	-45.0

All numbers are based on yearly returns. Return (%) is calculated geometrically. The Sharpe Ratio is calculated by subtracting the yearly risk-free rate from yearly returns, and then dividing by their yearly standard deviation. Turnover (TO) is measured by the sum of absolute deviations in weights before and after rebalancing. We rebalance portfolios every year, using past 250 trading days for our optimizations. VaR stands for Value-at-Risk, which indicates the minimum return (or equivalently, the maximum loss) for a certain significance level (95% or 99%). Short-selling is not allowed for all optimizations. All measures are calculated over the whole period of 1990 to 2010. Significant Sharpe Ratio and CEV3 differences of the alternative portfolio model and the Minimax at 1%, 5%, 10% level are denoted by ⋆⋆⋆, ⋆⋆ and ⋆, respectively.

3.A.2 Robustness Check: Daily Rebalancing

Table 3.A.5: Robustness Check: Daily Rebalancing

PERFORMANCE	Return (%)	Vola (%)	Sharpe Ratio	Excess Return	TO	CEV3
Minimax	0.04	0.44	0.058	over Minimax (%)	40.4	3.7
Minimum Variance	0.04	0.42	0.057	-0.002	41.8	3.6***
Equal Weights	0.04	0.79	0.036	0.003	0.0	3.4***
Fixed Weights	0.04	0.77	0.036	0.002	0.0	3.3**
Mean-Variance	0.06	0.94	0.052	0.023	394.7	4.9***

RISK	Volatility (%)	VaR(0.95) (%)	VaR(0.99) (%)	Max. Loss (%)
Minimax	0.44	-0.6	-1.0	-2.1
Minimum Variance	0.42	-0.5	-1.1	-2.6
Equal Weights	0.79	-0.9	-2.3	-8.7
Fixed Weights	0.77	-1.0	-2.1	-7.3
Mean-Variance	0.94	-1.7	-3.6	-16.8

All numbers are based on daily returns. Return (%) is calculated geometrically. The Sharpe Ratio is calculated by subtracting the daily risk-free rate from daily returns, and then dividing by their daily standard deviation. Turnover (TO) is measured by the sum of absolute deviations in weights before and after rebalancing. We rebalance portfolios every day, using past 250 trading days for our optimizations. CEV3 is the Certainty Equivalent for a Mean-Variance investor with risk aversion parameter $\lambda = 3$. VaR stands for Value-at-Risk, which indicates the minimum return (or equivalently, the maximum loss) for a certain significance level (95% or 99%). Short-selling is not allowed for all optimizations. All measures are calculated over the whole period of 1990 to 2010. Significant Sharpe Ratio and CEV3 differences of the alternative portfolio model and the Minimax at 1%, 5%, 10% level are denoted by $\star\star\star$, $\star\star$ and $\star$, respectively.

3.A.3 Robustness Check: Short-Selling allowed

Table 3.A.6: Robustness Check: Short-Selling allowed (Yearly Rebalancing)

PERFORMANCE	Return (%)	Vola (%)	Sharpe Ratio	Excess Return	TO	CEV3
Minimax	0.05	0.54	0.071	over Minimax (%)	+9.2	4.8
Minimum Variance	0.04	0.4	0.063	-0.013	10.2	3.8***
Mean-Variance	0.58	4.3	0.131	+0.53	243.3	39.2***

RISK	Volatility (%)	VaR(0.95) (%)	VaR(0.99) (%)	Max. Loss (%)
Minimax	0.54	-0.6	-1.0	-2.1
Minimum Variance	0.40	-0.5	-1.1	-2.6
Mean-Variance	4.30	-6.0	-10.7	-31.9

Table 3.A.7: Robustness Check: Short-Selling allowed (Daily Rebalancing)

PERFORMANCE	Return (%)	Vola (%)	Sharpe Ratio	Excess Return	TO	CEV3
Minimax	0.04	0.44	0.058	over Minimax (%)	40.4	3.7
Minimum Variance	0.04	0.57	0.045	0.0	46.5	3.6***
Mean-Variance	0.21	4.90	0.040	+0.17	3818.9	-2.9***

RISK	Volatility (%)	VaR(0.95) (%)	VaR(0.99) (%)	Max. Loss (%)
Minimax	0.44	-0.6	-1.0	-2.1
Minimum Variance	0.57	-0.6	-1.1	-14.7
Mean-Variance	4.90	-7.0	-12.7	-60.3

All numbers are based on daily returns. Return (%) is calculated geometrically. The Sharpe Ratio is calculated by subtracting the daily risk-free rate from daily returns, and then dividing by their daily standard deviation. Turnover (TO) is measured by the sum of absolute deviations in weights before and after rebalancing. We rebalance portfolios every year in Table 3.A.6 and every day in Table 3.A.7, using past 250 trading days for our optimizations. CEV3 is the Certainty Equivalent for a Mean-Variance investor with risk aversion parameter $\lambda = 3$. VaR stands for Value-at-Risk, which indicates the minimum return (or equivalently, the maximum loss) for a certain significance level (95% or 99%). Short-selling is allowed for all optimizations. All measures are calculated over the whole period of 1990 to 2010. Significant Sharpe Ratio and CEV3 differences of the alternative portfolio model and the Minimax at 1%, 5%, 10% level are denoted by $\star\star\star$, $\star\star$ and $\star$, respectively.

3.A.4 Robustness Check: Crises Periods

Table 3.A.8: Performance in Crises

	Internet bubble 2000/2001			Terrorist attack 9/11/2001			Banking crises 2007/2008			Recession 2008-2010		
	CE3	SR	Max Loss	CE3	SR	Max Loss	CE3	SR	Max Loss	CE3	SR	Max Loss
Minimax	8.1	0.104	1.71	3.6	0.045	1.50	3.9	0.073	1.05	4.6	0.076	2.06
MIV	7.7	0.107**	1.89	3.8	0.046	1.56	3.7	0.077***	1.31	4.5	0.074***	2.05
EW	5.9	0.059	5.06	2.7	0.031	1.62	2.3	0.024***	1.89	-2.9**	-0.006***	8.71
FW	2.6	0.029***	6.01	1.7	0.015***	1.78	2.6	0.029***	1.93	-1.7**	-0.004***	7.31
MV	8.5**	0.075**	4.48	-4.2	-0.016	3.20	10.8**	0.105***	3.01	7.0	0.075***	3.10

This table reports out-of-sample performance (CE3, SR, Max. Loss in %) for the Minimax portfolio and its competitors in several crises periods. We analyze the following crises: (1) Internet bubble 2000/2001, (2) Terrorist attack 9/11/2001, (3) Banking crises 2007/2008 and (4) Economic recession 2008-2010. Significant Sharpe Ratio and CEV3 differences of the alternative portfolio model and the Minimax at 1%, 5%, 10% level are denoted by $\star\star\star$, $\star\star$ and $\star$, respectively.

Bibliography

BELOUSOVA, J. AND G. DORFLEITNER (2012): "On the diversification benefits of commodities from the perspective of euro investors," *Journal of Banking and Finance*, 36, 2455–2472.

BIGLOVA, A., S. ORTOBELLI, S. RACHEV, AND S. STOYANOV (2004): "Different Approaches to Risk Estimation in Portfolio Theory," *The Journal of Portfolio Management*, 31 (1), 103–112.

BRAINARD, K. (2009): "Summary of Findings for FY 2008," *Public Fund Survey*, available from: www.publicfundsurvey.org, 10.

BRITTEN-JONES, M. (1999): "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights." *Journal of Finance*, 54(2), 655–671.

CAI, X., K. TEO, X. YANG, AND X. ZHOU (2000): "Portfolio optimization under a minimax rule," *Management Science*, 46 (7), 957–972.

CHATEAUNEUF, A., F. MACCHERONI, M. MARINACCI, AND J.-M. TALLON (2005): "Monotone continuous multiple priors," *Economic Theory*, 26 (4), 973–982.

DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): "Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms." *Management Science*, 55(5), 798–812.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22 (5), 1915–1953.

DING, Y. (2006): "Portfolio Selection under Maximum Minimum Criterion," *Quality and Quantity*, 40 (3), 457–468.

FAN, J., J. ZHANG, AND K. YU (2012): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios," *Journal of American Statistical Association*, 107, 592–606.

FROST, P. A. AND J. E. SAVARINO (1986): "An empirical Bayes approach to portfolio selection." *Journal of Financial and Quantitative Analysis*, 21, 293–305.

GARLAPPI, L., R. UPPAL, AND T. WANG (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *The Review of Financial Studies*, 20, 41–81.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18 (2), 141–153.

Hodder, J., J. Jackwerth, and O. Kolokolova (2012): "Improved Portfolio Choice Using Second-Order Stochastic Dominance," *Working Paper*.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps." *Journal of Finance*, 54(4), 1651–1684.

Jorion, P. (1986): "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21(3), 2279–292.

——— (1991): "Bayesian and CAPM estimators of the means: Implications for portfolio selection." *Journal of Banking and Finance*, 15, 717–727.

Kan, R. and G. Zhou (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *Journal of Financial and Quantitative Analysis*, 42(3), 621–656.

Ledoit, O. and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004a): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 21(1), 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

Lintner, J. and W. Krasker (1982): "Minimax behavior in portfolio selection," *Journal of Finance*, 37 (2), 609–614.

Mansini, R., W. Ogryczak, and G. Speranza (2003): "LP solvable models for portfolio optimization: a classification and computational comparison," *IMA Journal of Management Mathematics*, 14 (3), 187–220.

Markowitz, H. M. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

Merton, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics*, 8, 323–361.

Michaud, R. (1989): "The Markowitz optimization enigma: Is optimized optimal?" *Financial Analysts Journal*, 45, 31–42.

Sengupta, J. (1982): "A minimax policy for optimal portfolio choice," *International Journal of Systems Science*, 13 (1), 39–56.

Teo, K. and X. Yang (2001): "Portfolio Selection Problem with Minimax Type Risk Function," *Annals of Operations Research*, 101 (1), 333–349.

Tuetuencue, R. and M. Koenig (2004): "Robust Asset Allocation," *Annals of Operations Research*, 132, 157–187.

Young, M. (1998): "A Minimax Portfolio Selection Rule with Linear Programming Solution," *Management Science*, 44 (5), 673–683.

Chapter 4

Regularization of Portfolio Weights using the LASSO

4.1 Introduction

Portfolio management classes often teach the standard portfolio choice framework of Markowitz (1952). In practice, however, the mean-variance portfolio is difficult to implement. The optimal weights depend sensitively on the return μ and the covariance matrix Σ. Estimation errors of both input parameters can be large. It has been found that in particular expected returns are hard to estimate (Merton, 1980). Also the estimated covariance matrix may be close to singular (Chan et al., 1999). Estimation errors lead to a suboptimal portfolios with poor performance and extreme short positions (Jorion, 1991). Several attempts to stabilize portfolio weights have been made. Considering the first problem, many authors avoid the estimation of returns and concentrate on the minimum variance portfolio (see e.g. Frost and Savarino, 1986). Alternatively, shrinkage approaches to the mean have been proposed (Jorion, 1986). To solve the second problem concerning the singularity of the estimated covariance matrix, shrinkage approaches have been proposed (Ledoit and Wolf, 2003, 2004). Portfolios with no short-sale constraints (Jagannathan and Ma, 2003) or equal weighting (DeMiguel et al., 2009b) show a surprisingly good performance.

A recent approach to tackle portfolio instability imposes norm constraints (Brodie et al., 2009). Instead of shrinking the covariance matrix, the norm of the weights is restricted. Both approaches are mathematically identical (DeMiguel et al., 2009a). The restriction of portfolio weights seems to be more intuitive than the shrinkage of the covariance matrix. An investor might have a prior belief about the weights but less about the returns' covariance matrix. Moreover, the L_1 norm restriction of the weights can be used as a short-selling constraint. This type of restriction leads to a selection of assets. The higher the penalty of the weights' norm, the fewer assets are included. The question how the magnitude of the penalty should be chosen is still open. DeMiguel et al. (2009a) select the optimal penalty by cross-validation. However, the optimal penalization strongly varies over time and is difficult to forecast. It also implies an additional source of parameter uncertainty. Brodie et al. (2009) and Fan et al. (2012) fix the number of assets and the maximum L_1 norm of the weights, respectively. For both approaches, there exists no rule on how to choose the respective level. A stable solution for the penalization intensity is required. Additionally the solution should be easy to implement. Our contribution to literature is to propose an easy to calculate short-selling penalty level. The penalty term is to derive by a rule-of-thumb and is asymptotically optimal. The penalized portfolio proves

to perform well in a simulation study as well as in an empirical study compared to several well-known benchmark portfolio strategies.
The paper is organized as follows. In Section 2, the idea of portfolio regularization is reviewed. We derive some theoretical results for the CE loss. We find an asymptotically optimal level of regularization. In Section 3, we present a simulation study to determine how regularization performs for a known return process. In Section 4, the performance of the regularization approach is tested against alternative models using the Fama-French industry data sets for different number of assets. Finally, Section 5 concludes.

4.2 Portfolio Optimization

4.2.1 Setup

Consider a market with n assets. The (excess) return of the assets at time t is given by the $n-$dimensional return vector $\mathbf{r}_t = (r_{1,t}, \ldots, r_{n,t})'$. For simplicity we drop time index t. The first two moments of the return vector is given by $E(\mathbf{r}) = \mu \in \mathbb{R}^n$ and $V(\mathbf{r}) = \Sigma \in \mathbb{R}^{n\times n}$. A portfolio is defined by a weight vector $w = (w_1, \ldots, w_n)' \in \mathbb{R}^n$ with $w'\iota = 1$ and ι being a vector of ones. Weight w_i denotes the share invested in asset i. Hence, the investor is fully invested in all n assets, including some potential short positions. The mean and variance of the portfolio are given by $w'\mu$ and $w'\Sigma w$. The performance of a portfolio is evaluated by the Certainty Equivalent (CE) which is defined by

$$CE(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w$$

with γ being the risk parameter. The CE covers the risk-neutral investor ($\gamma = 0$) as well as the minimum variance investor ($\gamma \to \infty$). The CE is equivalent to the Markowitz (1952) portfolio optimization problem. It can be shown that the CE reflects the investor's loss function if (i) his utility function is quadratic or (ii) his utility function is exponential and returns $\mathbf{r}$ are normally distributed or (iii) the investment horizon is short (see e.g. Schanbacher, 2012).
The mean returns and covariances are unknown. Their estimates are often imprecise and lead to unstable and extreme portfolio weights (Best and Grauer, 1991b). We follow a regularization approach for the norm of the weights to counteract extreme portfolio weights.

4.2.2 L_q Regularization

Instead of many attempts to shrink the mean or the covariance[1], our regularization approach introduces a penalty on the weights directly. The approach is more intuitive as the investor is not directly interested in the structure of the returns' moments. His goal is to have stable and less extreme weights. The corresponding optimization function is given by

$$CE_\lambda(w) = w'\mu - \frac{\gamma}{2} w'\Sigma w - \lambda ||w||_q \tag{4.1}$$

The optimization reflects the common CE with an additional penalty term for the weight. A large penalty level $\lambda \geq 0$ reflects the investor's preference for weights with a smaller L_q−norm[2]. The optimal weights maximize the CE (eq. 4.1) and are denoted by

$$w^\star_\lambda = \arg \max_{w:w'\iota=1} w'\mu - \frac{\gamma}{2} w'\Sigma w - \lambda ||w||_q \tag{4.2}$$

The impact of the penalty term is shown by the following propositions. It shows that a higher penalty parameter λ leads to an optimal portfolio with a smaller L_q norm.

Proposition 4.2.1. For $\underset{\sim}{\lambda} \leq \tilde{\lambda}$ it follows that $||w^\star_{\underset{\sim}{\lambda}}||_q \geq ||w^\star_{\tilde{\lambda}}||_q$.

Proof. The proof is given in 4.A.1 □

We find that increasing the penalty factor decreases the norm of the optimal weights. As the weights $w \in \mathbb{R}^n$ should sum up to one, there exists a minimum barrier the optimal weights can converge to. The target depends on the L_q norm. We consider the two most important special cases. If a L_1 penalty term (i.e. $q = 1$ in eq. 4.2) is used, it is called the Lasso (Tibshirani, 1996). The most prominent alternative is the ridge regression (Tychonoff and Arsenin, 1977) based on the L_2 penalty (i.e. $q = 2$ in eq. 4.2). For those two cases, the following proposition shows the optimal weights for a high penalty level.

Proposition 4.2.2. Consider the optimal weights $w^\star_\lambda$ (eq. *4.2*) with increasing penalty level, i.e. $\lambda \to \infty$. Then $w^\star_\lambda$ converges to

- the no short-sale portfolio $w_{MVNS} = \arg\max_{w:w'\iota=1, w_i \geq 0} CE(w)$ if the Lasso is applied (i.e. $q = 1$).

[1]See e.g. Jorion (1986) for shrinking the mean and Ledoit and Wolf (2003) for shrinking the covariance matrix.

[2]The L_q norm is defined by $||w||_q = \sum_{i=1}^N |w_i|^q$

- equally weighted portfolio $w_{EQ} = (\frac{1}{n}, \ldots, \frac{1}{n})'$ if the ridge regression is used (i.e. $q = 2$).

Proof. The proof is given in 4.A.2 □

We observe that the characteristics of the optimal weights are distinct for different norms. The no short-sale portfolio is known to be sparse. Only few assets are included into the portfolio, i.e. $\#\{w_i \in w^\star_\lambda : w_i \neq 0\} < n$. Using the L_1 penalty, one obtains a portfolio with low diversification but a selection of the driving assets. The opposite holds for the L_2 penalty. By definition the equally weighted portfolio has a positive weight on each asset. The equally weighted portfolio maximizes diversification but is data-ignorant.
Under the L_2 penalty a closed form solution for the optimal weights exists. The ridge regression can be interpreted as a shrinkage approach to the identity matrix.

$$\begin{aligned} CE_\lambda(w) &= w'\mu - \frac{\gamma}{2} w'\Sigma w - \lambda ||w||_2 \\ &= w'\mu - \frac{\gamma}{2} w'\Sigma w - \lambda w' I w \\ &= w'\mu - \frac{\gamma}{2} w' \underbrace{(\Sigma + \frac{2\lambda}{\gamma} I)}_{\tilde{\Sigma}} w \end{aligned}$$

Instead of the actual covariance matrix, the investor optimizes with respect to $\tilde{\Sigma} = \Sigma + \frac{2\lambda}{\gamma} I$. The penalty λ transforms into the shrinkage intensity for the covariance matrix. The optimal weights are then given by (see e.g. Best and Grauer, 1991a)

$$w^\star_\lambda = \frac{\tilde{\Sigma}^{-1}\iota}{\iota'\tilde{\Sigma}^{-1}\iota} + \frac{1}{\gamma}\left(\tilde{\Sigma}^{-1} - \frac{\tilde{\Sigma}^{-1}\iota\iota'\tilde{\Sigma}^{-1}}{\iota'\tilde{\Sigma}^{-1}\iota}\right)\mu$$

It is possible to determine an optimal level of λ. The L_2 penalty leads to a solution which gives a certain (generally positive) weight to each asset. This is theoretically an advantage as the portfolio is well-diversified. Practically, it is an disadvantage as the transaction and monitoring costs are high. Often an investor is interested in a sparse portfolio which consists of a limited number of assets. This investor might be interested in the L_1 penalty to select relevant assets. Unfortunately for the L_1 penalty, a closed form solution does not exist. Efron et al. (2004) present a numerical algorithm to solve the Lasso using the least-angle regression (LARS). Our problem in equation (4.2) is no standard regression problem as it also includes the constraint that weights have to sum up to one, i.e. $w'\iota = 1$. For our application we use the iterative shooting algorithm of Loris and Verhoeven (2012). A brief description can be found in 4.B.1. The algorithm finds the solution path of $w^\star_\lambda$ for all $\lambda \geq 0$. For a large penalty λ the optimal weights $w^\star_\lambda$ correspond to the weights of the no short-sale portfolio. Generally, this portfolio is the sparsest portfolio obtained by the Lasso. As λ decreases the algorithm successively recruits more assets. With $\lambda = 0$ the optimal weights $w^\star_\lambda$ are the Markowitz weights of the unconstrained portfolio. Commonly, in the Markowitz case all assets have a non-zero weight and many assets are held short. The core question is how to determine the penalty term.

4.2.3 L_1 Regularization

The findings of Proposition 4.2.2 should make us sensitive how different L_q norms can result in portfolios with completely different characteristics. Suppose that the investor considers transaction and monitoring costs. She wants to select a limited number of assets. From now on, we restrict our analysis on the Lasso case (i.e. $q = 1$). The L_1 norm can be interpreted as the amount of short positions.

$$\begin{aligned} ||w||_1 &= \sum_i |w_i| \\ &= \sum_i w_i - 2 \cdot \sum_{i:w_i<0} w_i \\ &= 1 + 2 \cdot \text{Short Positions} \end{aligned}$$

Penalizing the L_1 norm is equivalent of penalizing short positions. Fan et al. (2012) fix the number of allowed short positions. Restricting $||w||_1 \leq 1 + 2 \cdot c$ allows for a share c of short positions. The drawback of this approach is, that in some cases it might be reasonable to buy more short positions than allowed by the fixed level c. For instance, if a large fraction of the portfolio can be or should be hedged, one needs to hold large short-positions. In this case, a strict binding limit is harmful to the portfolio's performance. Further, it is unclear on how to determine c. Fan et al. (2012) select different levels c without advising a particular value. Instead of fixing c, Brodie et al. (2009) select the number of assets which should be included in the portfolio. Penalty level λ is chosen such that $k = \#\{w_i \in w^\star_\lambda : \ 0 \neq w_i\}$ for some target value k. The drawback of this approach is that there might be no solution. If the no short-sale portfolio contains more assets than the target value, i.e. $\#\{w_i \in w^\star_\infty : \ 0 \neq w_i\} > k$, there is no penalty λ satisfying the objective function. Brodie et al. (2009) also do not advice a specific number of assets. They find that "the no-short-positions portfolio outperforms all binned portfolios for the full 30-year period" (Brodie et al., 2009, p. 12271). Their approach of selecting the number of assets k seems to be difficult if the optimum is close to the no short-sale portfolio. The investor has to select k to be small to get close to the no short-sale portfolio. At the same time, he runs into the risk of choosing k too small to obtain a solution at all.

Instead of fixing the amount of short-selling c or the number of assets k, we include the penalty term for the L_1 norm of the weights, i.e. $\lambda||w||_1$. The approach is mathematically equivalent to both approaches for some function $c(\lambda) : \lambda \mapsto c$ or $k(\lambda) : \lambda \mapsto k$. A fixed penalty λ can be imitated by a moving norm barrier c or a flexible target number of assets k. In common applications one either fixes penalty λ, the admissible amount of short-selling c or the maximum number of assets k. Depending on the situation, there may be good reasons for extreme short positions or a high number of assets. Instead of introducing a strict barriers for c or k, it seems economically more sensible to apply a penalty which allows for extreme weights, if potential benefits are high (compare to eq. 4.2).

Before we analyze the estimation risk, we first analyze why it might be reasonable to discard weights with too extreme L_1 norms. The following proposition gives an intuition that the optimal weights should not have an L_1 norm being too large.

Proposition 4.2.3. *The optimal weights $w^\star$ have no short positions (i.e. $||w^\star||_1 = 1$) if*

a.) the expected returns are non-negative, i.e. $\mu_i \geq 0\ \forall i$, and the returns are only moderately positive correlated.

or

b.) the expected returns are of similar magnitude, i.e. $\mu_i = \underline{\mu} \in \mathbb{R}\ \forall i$, and the returns are only positively correlated due to a common risk factor, i.e. $\sigma_i^2 = \tilde{\sigma}_i^2 + \sigma_\eta^2$ where σ_η^2 represents the systematic risk of the common risk factor.

Proof. The proof is relegated to 4.A.3 □

Proposition 4.2.3 gives an intuition that the L_1 norm of the true weights should not be too large. The estimated weights, however, tend to be of extreme magnitudes. Suppose that the means are identical and returns are highly correlated due to a common risk factor. In finite samples the estimated means are unequal. Together with an almost singular covariance matrix, the estimated weights are of extreme magnitudes. In light of Proposition 4.2.3, most likely an estimation error is responsible for an extreme L_1 norm of estimated portfolio weights. A penalization of the L_1 norm seems to be reasonable.

4.2.4 Estimation Risk

So far we considered the effect of a penalty term given μ and Σ. In practice, the moments have to be estimated. Empirically, the weights of the penalized objective function often perform better than the theoretical superior weights based on the unpenalized objective function. Regularization reduces estimation risk at the cost of a small bias. The estimates of the returns' first two moments are denoted by $\hat{\mu}$ and $\hat{\Sigma}$. Then the estimated counterpart of the penalized CE in equation (4.1) is given by

$$\hat{CE}_\lambda(w) = w'\hat{\mu} - \frac{\gamma}{2}w'\hat{\Sigma}w - \lambda||w||_1$$

The estimated optimal weights are given by

$$\hat{w}_\lambda = \arg\max_{w:w'\iota=1} w'\hat{\mu} - \frac{\gamma}{2}w'\hat{\Sigma}w - \lambda||w||_1 \tag{4.3}$$

Henceforth, $\hat{CE}$ and $\hat{w}$ denote the unrestricted case (i.e. $\lambda = 0$). When is the performance of estimated weights considerably bad? The following proposition introduces an upper bound of the performance difference of the theoretical optimal weight and the estimated optimal weight. The results generalize the findings of Fan et al. (2012).

Proposition 4.2.4. *Let $w^\star$ be the theoretical optimal weight and $\hat{w}$ its estimated counterpart. Then the following upper bound to the performance difference in terms of the CE exists:*

$$0 \leq CE(w^\star) - CE(\hat{w}) \leq 2 \cdot \max_{w \in \{w^\star, \hat{w}\}} |CE_\lambda(w) - \hat{CE}_\lambda(w)| \tag{4.4}$$

Proof. The proof is shifted to 4.A.4 □

Equation (4.4) shows that the difference in performance (i.e. CE) can be large if the difference of the estimated and the true penalized performance (i.e. CE_λ) is large. The following proposition connects estimation errors in the moments with the L_1 penalty of the weights.

Proposition 4.2.5. *For any weight w it holds that*

$$|CE_\lambda(w) - \hat{CE}_\lambda(w)| \leq ||\mu - \hat{\mu}||_\infty ||w||_1 + \frac{\gamma}{2}||\Sigma - \hat{\Sigma}||_\infty ||w||_1^2 \tag{4.5}$$

with $\hat{\mu}$ *and* $\hat{\Sigma}$ *being some estimates of* μ *and* Σ*. The maximum norm is denoted by* $||A||_\infty = \max\{|A_{1,1}|, \ldots, |A_{m,n}|\}$*. For the special case that* $\hat{\mu} \in \mathbb{R}^n$ *and* $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ *are the sample mean and the sample covariance matrix over* T *periods, it holds that*

$$||\mu - \hat{\mu}||_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log n}{T}}\right) \tag{4.6}$$

$$||\Sigma - \hat{\Sigma}||_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log n}{T}}\right) \tag{4.7}$$

Proof. The proof is shifted to 4.A.5 □

With Propositions 4.2.4 and 4.2.5, we find that the performance loss has a lower bound if $||\mu - \hat{\mu}||_\infty$ and $||\Sigma - \hat{\Sigma}||_\infty$ are small. Equation (4.5) shows that the estimation risk does not accumulate. The covariance matrix can become very large. For 100 available assets, one has to estimate 5050 parameters of the covariance matrix. Aggregating of estimation error in each of the thousands of parameters could lead to an enormous CE loss. The result suggests that only the single worst element-wise estimation error determines the CE difference (see also Fan et al., 2012). Given a moderate elementwise estimation error of the estimated moments, the CE loss depends crucially on the L_1 norm of the weight (see eq. 4.5). The unconstrained Markowitz weights tend to have an extremly large L_1 norm. A huge CE loss is possible. A regularization of the L_1 norm reduces the norm of the weights; thereby, the maximum possible CE loss.
Now we consider the special case of using the sample mean and sample covariance as the estimates of the returns' first two moments. The estimation error of the moments decreases as $\frac{\log n}{T}$ decreases. For a long estimation window T or a small portfolio size n, the estimation risk in the sample estimates is small. Note that the portfolio size n enters at logarithmic order only while the estimation period T enters linearly. The convergence result shows that for large $\log(n)$ compared to estimation window T the performance difference of true and estimated weights can be high (see proposition 4.2.4). In this case a high level of regularization is desirable.
The convergence rate of the sample estimates of μ and Σ can be calculated (eq. 4.6 and 4.7). Unfortunately finite sample results cannot be obtained. Strictly speaking, the limit does not give any advice for finite sample properties. It gives an idea how the penalty parameter should be chosen asymptotically. For large $\frac{\log n}{T}$ high regularization is needed and for small $\frac{\log n}{T}$ we do not need to include a penalty term.

We propose to select

$$\lambda = c \cdot \frac{\log n}{T} \tag{4.8}$$

with some scaling constant $c > 0$. Scaling factor c has to be determined by a simulation study. The performance of the proposed penalty of eq. 4.8 has to be analyzed empirically. The objective function is then given by

$$\hat{w} = \arg \max_{w:w'\iota=1} w'\hat{\mu} - \frac{\gamma}{2} w'\hat{\Sigma} w - c \cdot \frac{\log n}{T} ||w||_1 \tag{4.9}$$

4.2.5 Extensions: Portfolio Tracking and Hedging

Several portfolio strategies in practice are closely connected to the regularization problem described above. One of the main extensions is index tracking, which is crucial for the growing market of ETFs (Exchange-Traded Funds). The investor wants to track an index with a large number of assets. Buying every single underlying asset will exactly mimic the index but largely increases trading and monitoring costs. The investor might prefer to choose a subset of assets, which closely tracks the index level but still keeps transaction costs low. The appropriate set of assets can be selected using an L_1-penalty for the weights.
The objective function can be rewritten into

$$\hat{w} = \arg \min_{w:w'\iota=1} ||Rw - y||_2^2 + \lambda ||w||_1 \tag{4.10}$$

for return matrix $R \in \mathbb{R}^{n \times T}$ and target return of the index $y \in \mathbb{R}^T$ (see 4.B.2). Without the penalty (i.e. $\lambda = 0$) the investor chooses the weights, which are closest to the index in the L_2 sense. The penalty term $\lambda||w||_1$ helps to select a subset of relevant assets. The penalty level λ can be chosen as given in eq. 4.8. Alternatively, the investor defines a maximum number of assets she is willing to buy and monitor. In this case he can set the penalty λ accordingly.
The same procedure can be also directly applied to portfolio hedging. Assume a given risky portfolio cannot be eliminated for reasons such as liquidity constraints, legal restrictions, costumer's obligations, or others. The short position of the portfolio might be too expensive. The cost-effective solution is to track the short position by only a small subset of assets. Relevant assets can be selected similar to the portfolio tracking approach of (4.10).

4.3 Simulation Study

The advantage of the simulation study is that the true mean and covariance are known. The theoretical optimal weights and the optimal CE can be calculated. The study analyzes the finite sample behavior of our proposed penalty rule $\lambda = c \cdot \log n/T$ with respect to varying size of the portfolio and of the estimation window. By the simulation also the scaling factor c in (4.8) can be determined.

4.3.1 Simulation Procedure

A simulation method to generate a wide range of distributional parameters is needed. We use the simulation settings of Kempf and Memmel (2006). The simulation procedure is as follows:

1. Fix the number of assets n and the length of the estimation window T.
2. In replication $s \in \{1, \ldots, S\}$, draw the moments of the return, Σ_s and μ_s of an inverse Wishart distribution and a conditional normal, resp.

   $$\Sigma_s \sim W^{-1}\left((n\Sigma_{hyper})^{-1}, k, n\right)$$

   and

   $$\mu_s|\Sigma_s \sim N\left(\mu_{hyper}\iota, \Sigma_s/\tau\right)$$

   The hyperparameter of the mean and covariance as well as k, τ are defined as in Kempf and Memmel (2006). However, we find that the results are robust to changes of the hyperparameters. The simulation method covers a wide area of possible distributional parameters.
3. Simulate a return series $\{r_t^s\}_{t=1}^T$ based on the generated parameters μ_s, Σ_s;

   $$r_t^s \overset{iid}{\sim} N(\mu_s, \Sigma_s) \quad \text{for } t = 1, \ldots, T$$

   with r_t^s being a vector of n returns at time t and replication s.
4. Estimate the sample mean and variance $\hat{\mu}_s, \hat{\Sigma}_s$, based on the return series $\{r_t^s\}_{t=1}^T$.
5. Determine the optimal weight $\hat{w}_\lambda^s$ for $\lambda \geq 0$ according to equation (4.3),

   $$\hat{w}_\lambda^s = \arg\max_{w:w'\iota=1} w'\hat{\mu}_s - \frac{\gamma}{2} w'\hat{\Sigma}_s w - \lambda||w||_1$$

We report the results for risk aversion level $\gamma = 2$. In 4.D.1 we provide a robustness check to show that the results are robust to changes in the level of risk aversion γ.

6. As the mean μ_s and variance Σ_s are known, the true CE of $\hat{w}_\lambda^s$ can be calculated. The optimal penalty level is then given by

$$\begin{aligned}\lambda_s^\star &= \arg\max_{\lambda \geq 0} CE(\hat{w}_\lambda^s) \\ &= \arg\max_{\lambda \geq 0} \hat{w}_\lambda^{s\prime}\mu - \frac{\gamma}{2}\hat{w}_\lambda^{s\prime}\Sigma\hat{w}_\lambda^s\end{aligned}$$

7. Repeat the steps (2)-(6) S times. For each replication $s \in \{1, \ldots, S\}$, we obtain an optimal penalty level $\lambda_s^\star$.

8. The optimal penalty level $\lambda^\star$ is determined by the average over all replications s, i.e. $\lambda^\star = \frac{1}{S}\sum_{s=1}^{S} \lambda_s^\star$. We find, that the optimal penalty level $\lambda^\star$ is robust estimated by $S = 1000$ replications.

9. The procedure is repeated for different number of assets n and different estimation windows T. The optimal penalty level depends on those parameters, i.e. $\lambda^\star(n, T)$.

4.3.2 Simulation Results

We apply the simulation procedure described in Section 4.3.1 for different numbers of assets n and different estimation windows T. As common numbers of assets we choose $n \in \{6, \ldots, 48\}$ which covers the standard Fama-French data sets as well as common indices such as the Dow Jones Industrial. As the length of the estimation window we choose $T \in \{60, 70, 80, 90, 100, 110, 120\}$. For shorter horizons, e.g. $T = 30$, the covariance matrix becomes singular if $n > 30$. We obtain $43 \cdot 7 = 301$ cases covering a small portfolio as well as a large portfolio for medium to large estimation windows. In the 4.C.1 the CE of $\hat{w}_\lambda$ is given for $\lambda \in [0, 0.025]$. The λ leading to the highest CE refers to the optimal penalty term $\lambda^\star$. We find that the optimal penalty is larger if the portfolio is large (i.e. n large) or the estimation window is short (i.e. T small). In these cases the no short-sale portfolio performs almost as good as the optimal penalized portfolio. These findings are in line with our theory. If the estimation error is high, a high level of regularization is needed. The no short-sale portfolio represents the strongest regularization in our framework and can be

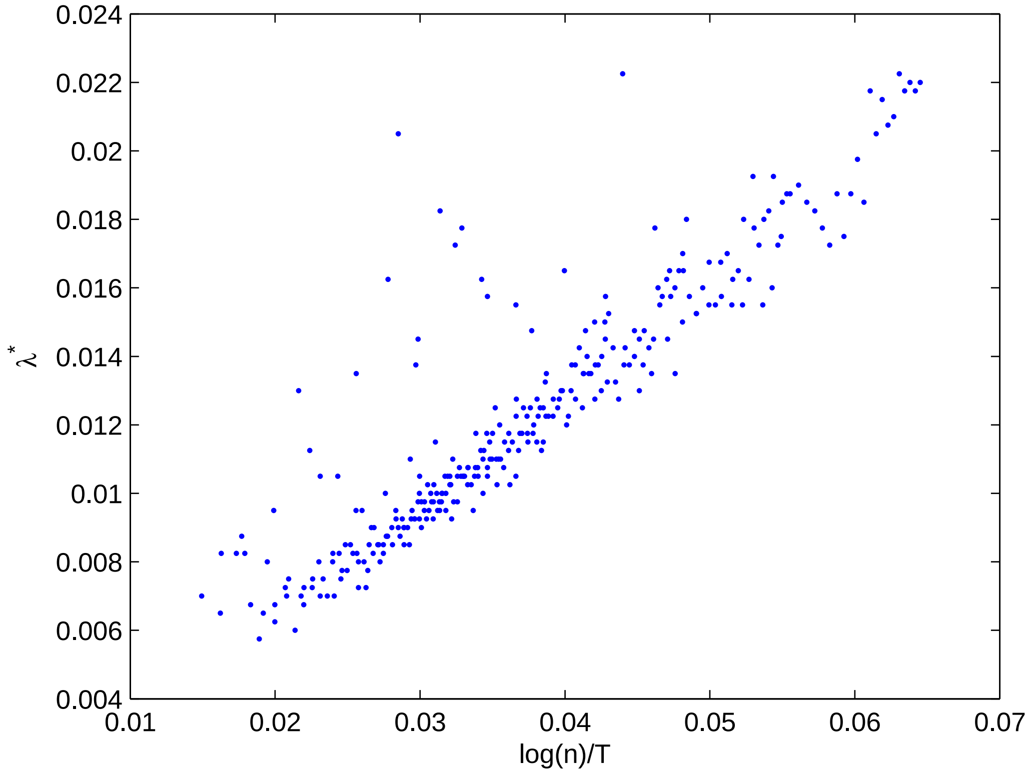


Figure 4.1: Regression of $\lambda^\star$ on $\log(n)/T$
The optimal penalty term $\lambda^\star$ ($y-$axis) is determined for $n \in \{6, \ldots, 48\}$ and $T \in \{60, 70, 80, 90, 100, 110, 120\}$. The penalty is plotted against $\log(n)/T$ ($x-$axis).

optimal in presence of high estimation risk.
Consider the relationship of the optimal penalty to the number of assets and the estimation window. A scatter plot of the optimal penalties $\lambda^\star$ and the corresponding value $\log n/T$ is provided in Figure 4.1. We obtain a clear linear relationship which is confirmed by the regression output in Table 4.1. At a 1% level, the constant is not significantly different from zero. The slope is given by a strongly significant value of about 0.3. We find that the following relationships approximately holds for all analyzed assets and rolling windows:

$$\lambda^\star \approx 0.3 \cdot \frac{\log n}{T} \tag{4.11}$$

We refer to Equation (4.11) as the rule of thumb for the penalty term. For any data set and estimation window, one can directly commute an approximate optimal penalty level $\lambda^\star$. In the following, we analyze the performance of the rule of thumb based on an empirical exercise.

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0004	0.0002	1.9858	0.0480
log(n)/T	0.3037	0.0051	59.6132	0.0000
R-squared	0.9224	Mean dependent var		0.0116
Adjusted R-squared	0.9221	S.D. dependent var		0.0035
S.E. of regression	0.0010	Akaike info criterion		-11.0200
Sum squared resid	0.0003	Schwarz criterion		-10.9954
Log likelihood	1660.5	F-statistic		3553.7
Durbin-Watson stat	1.4413	Prob(F-statistic)		0.0000

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
log(n)/T	0.3134	0.0015	214.1752	0.0000
R-squared	0.9214	Mean dependent var		0.0116
Adjusted R-squared	0.9214	S.D. dependent var		0.0035
S.E. of regression	0.0010	Akaike info criterion		-11.0135
Sum squared resid	0.0003	Schwarz criterion		-11.0012
Log likelihood	1658.5	Durbin-Watson stat		1.4487

Table 4.1: Regression of $\lambda^\star$ on $\log(n)/T$ with and without a constant.

4.4 Empirical Study

The rule of thumb of the penalty term is given by equation (4.11). In our simulation study we find that regularization of portfolio weights seems to be a promising strategy to improve the performance of the portfolio. In the following we conduct an empirical study to the regularization. The performance of the L_1 penalized portfolio is tested on the Fama-French industry data sets.

4.4.1 Procedure

The given data sets consist of T^{max} monthly returns of n assets, i.e. $r_t \in \mathbb{R}^n$ for $t = 1, \ldots, T^{max}$. Let the length of the estimation window be T. The optimal penalty is then given by $\lambda^\star = 0.3 \cdot \frac{\log n}{T}$. For each time t we estimate the optimal weights $\hat{w}_{\lambda^\star,t}$ based on the past T returns, i.e. $\mathcal{F}_t = \{r_{t-1}, \ldots, r_{t-T}\}$. The weights generate the out-of-sample return $\tilde{r}_t = \hat{w}'_{\lambda^\star,t} r_t \in \mathbb{R}$. The estimated out-of-sample performance of the model is then given by

$$\hat{CE} = \hat{\mu} - \frac{\gamma}{2}\hat{\sigma}^2$$

with $\hat{\mu} = \frac{1}{T^\star - T}\sum_{t=T+1}^{T^\star} \tilde{r}_t$ and $\hat{\sigma}^2 = \frac{1}{T^\star - T - 1}\sum_{t=T+1}^{T^\star}(\tilde{r}_t - \hat{\mu})^2$. As before we choose $\gamma = 2$. For robustness purpose we also considered other risk aversion levels. As expected by the robustness check of the simulation study (4.D.1), we find that the results are similar and are therefore are not stated. We want to improve the performance of the unrestricted mean variance approach. Additionally, we also compare the performance of our regularization method to common alternative models suggested by literature. In particular we apply the procedure to the following portfolio selection models:

1.) *L1 penalized portfolio (Lasso).* The weights are determined according to equation (4.9) with $\lambda = 0.3 \cdot \log n/T$.

2.) *Mean-Variance portfolio (MV).* The weights are determined by equation (4.9) with $\lambda = 0$. This portfolio is equivalent to the Markowitz (1952) portfolio.

3.) *No short-sale portfolio (MVNS).* The weights are determined by equation (4.9) with λ such that $w_i \geq 0$ f.a. $i \in \{1, \ldots, n\}$. Jagannathan and Ma (2003) show why the no short-sale portfolio can perform well, despite being theoretically inferior to its unrestricted counterpart.

4.) *Minimum variance portfolio (MinVar).* The weights are determined by equation (4.9) with $\lambda = 0$ and the restriction that $\mu_i = \mu_j$ f.a. $i, j \in \{1, \ldots, n\}$. As the error in the mean is in particular high, Merton (1980) argues for the Minimum variance

Data set	n	Time Period	Abbreviation
5 Industry Portfolio	5	07/1963-12/2011	Ind5
10 Industry Portfolio	10	07/1963-12/2011	Ind10
12 Industry Portfolio	12	07/1963-12/2011	Ind12
17 Industry Portfolio	17	07/1963-12/2011	Ind17
30 Industry Portfolio	30	07/1963-12/2011	Ind30
38 Industry Portfolio	38	03/2000-12/2011	Ind38
48 Industry Portfolio	48	07/1969-12/2011	Ind48

Table 4.2: List of data sets
The data sets considered in the empirical study. Each data set contains monthly returns of n assets in the giving time period. All data sets can be downloaded from Ken French's website. The data sets are described in Section 4.4.2.

portfolio.
5.) *Equally weighted portfolio (EQ)*. The weights are given by $w_i = \frac{1}{n}$ f.a. $i \in \{1, \ldots, n\}$. DeMiguel et al. (2009b) analyze the performance of the naive portfolio.
6.) *Ledoit and Wolf portfolio (LW)*. The Ledoit and Wolf portfolio is based on shrinking the covariance matrix $\hat{\Sigma}_{LW} = \delta F + (1-\delta)\hat{\Sigma}$. The target covariance matrix F and shrinkage itensity δ are defined as proposed by Ledoit and Wolf (2004).

4.4.2 Data Sets

We compare the out-of-sample performance of the portfolio mentioned in Section 4.4.1 relative to the L_1 penalized portfolio. As a performance criterion we use the CE. The horse-race is based on monthly returns of the Fama-French industry portfolios. The data sets are publically available at French's Webside[3]. We are interested how our rule-of-thumb performs for various asset classes. We use the all given industry portfolios with $n = 7, 10, 12, 17, 30, 38$ and 48 assets. The portfolios are denoted by $Ind\#n$ with $\#n \in \{7, 10, 12, 17, 30, 38, 48\}$. The time span considered ranges from 07/1963 to 12/2011. If the data set included missing values, we shortened the time period accordingly. Thereby, the period length varies across different data sets. The rolling window size is given by $T = 60$. Alternative values of T lead to similar results and are therefore not stated. Based on the past T observations, each month the portfolio weights are calculated for the following month. Table 4.2 presents an overview of the applied data sets, as well as further information to the particular

[3]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Model \ Ind	5	10	12	17	30	38	48
Lasso	0.0069	0.0144	0.0149	0.0074	0.0153	0.0028	0.0211
MV	-0.0436	-0.1963	-0.2702	-0.3943	-3.4092	-7.891	-119.0659
	(0.0055)	(0.0005)	(0.0025)	(0.0005)	(0.0005)	(0.028)	(0.0045)
MVNS	0.0068	0.0143	0.0144	0.0066	0.0143	0.0026	0.0188
	(0.4121)	(0.4895)	(0.4351)	(0.1908)	(0.2043)	(0.1733)	(0.0754)
MinVar	0.0046	0.0098	0.0098	0.0052	0.0095	-0.0007	0.0141
	(0.2752)	(0.2323)	(0.1768)	(0.2892)	(0.1054)	(0.3307)	(0.1374)
EQ	0.0077	0.015	0.015	0.0077	0.015	0.0031	0.0184
	(0.7008)	(0.544)	(0.51)	(0.5619)	(0.4481)	(0.5165)	(0.2323)
LW	0.0082	0.0122	0.012	0.008	0.0124	0.0007	0.0131
	(0.6978)	(0.3586)	(0.2912)	(0.5804)	(0.2542)	(0.3836)	(0.0599)

Table 4.3: CE for empirical data
The table presents the empirical CE of the models presented in Section 4.4.1 applied to the data sets presented in table 4.2. In parentheses are the $p-$values of the difference between the CE of each model from that of the L_1 penalized benchmark. The $p-$values are calculated by a bootstrap method proposed by Ledoit and Wolf (2008).

data set.

4.4.3 Results

According to the procedure described in Section 4.4.1 we calculate the out-of-sample returns for each model considered. The procedure is applied to all Fama-French industry data sets. Our proposed Lasso method serves as a benchmark model. The $p-$values are stated in parentheses and test for CE difference to the benchmark. The $p-$values are calculated by the bootstrap procedure of Ledoit and Wolf (2008, 2011) with minor adjustments to the CE (for the adjustment see also Schanbacher, 2012).
We estimate the best penalty level $\hat{\lambda}_t$ based on the past T out-of-sample returns. That is, which penalty level $\hat{\lambda}_t$ leads to the best out-of-sample CE in the previous period of length T. The optimal penalty for each point in time t is given in 4.C.2. We find that the optimal penalty level is unstable. Huge jumps from one period to the next make the penalty $\hat{\lambda}_t$ difficult to forecast. We consider a random walk of the penalty and chose $\hat{\lambda}_{t-1}$ to estimate the weights for time t. In this case the out-of-sample performance turns out to be poor (not stated). We conclude that it

is difficult to estimate or even forecast the optimal penalty level as it appears to be unstable. The finding supports the desire for a time stable penalty level which is easy to calculate. Note that in all cases the penalty $\lambda^\star$ based on our rule-of-thumb is close to the average penalty $\hat{\lambda}_t$. It seems that the rule-of-thumb leads to a reasonable level of penalization.

The empirical performance of all models can be found in Table 4.3. We find that the Lasso approach performs in all cases better than the Mean Variance (MV) model. The MV portfolio has extreme weights leading to volatile out-of-sample returns. It is not surprising to find a low CE for the unconstraint MV model. The direct competitor is the no short-sale portfolio (MVNS). The MVNS portfolio corresponds to the Lasso model using a high penalty. For all data sets, the Lasso outperforms the MVNS portfolio. A common alternative is the Minimum Variance portfolio (MinVar). The Lasso approach outperforms the MinVar for each industry data set. In line with DeMiguel et al. (2009b), we find that the equally weighted portfolio (EQ) performs well, despite its data ignorance. Overall the equally weighted portfolio shows a similar performance as the Lasso approach. Finally, the Ledoit-Wolf portfolio outperforms the Lasso in two data sets and is outperformed by the Lasso in the remaining 5 data sets.

Unfortunately only few models can be beaten significantly. This is in line with the common finding that significant performance differences of stable models are rare (DeMiguel et al., 2009b). Still, at a 10% significance level, the Lasso significantly outperforms other models 9 times, while it is never significantly outperformed. We conclude that the Lasso penalization based on the rule of thumb performs well. The Lasso performs in particular well compared to its restricted relative - the no short-sale portfolio.

4.5 Conclusion

The vast literature on estimation risk of portfolio choice shows the demand for regularization approaches. So far literature mainly focused on the improvement of estimating the returns' moments. The investor is more interested in the weights itself. We apply a regularization approach by introducing a L_1-penalty term for the weights. The penalty parameter has to be selected. Mainly three approaches have been suggested. DeMiguel et al. (2009a) estimate the penalty term by cross validation (Efron and Gong, 1983). Brodie et al. (2009) fix the number of non-zero weights. Fan et al. (2012) fix the L_1-norm of the weights on a predefined level. We find that the optimal penalty level is shaky and therefore hard to forecast. The final two approaches do not provide a guideline on how many assets should be included or how large the maximum norm of the weights should be. We provide theoretical results how the CE loss behaves asymptotically with respect to the number of assets and the size of the rolling window. In a simulation study we find that the asymptotic results still hold in finite samples. We present a rule-of-thumb for the penalty term. In an empirical study we show that the L_1 regularization using our rule of thumb performs well compared to common alternative portfolio models. Practitioners can benefit of our simple and easy to apply penalty level. Academics have a benchmark for more advanced regularization methods. We believe that the potential to counteract estimation risk in portfolio optimization is far from being exploited.

Appendix 4.A Proofs

4.A.1 Proof to Proposition 4.2.1

Proof. Consider the CE of eq. 4.1 with $\underset{\sim}{\lambda}$. Further, consider the following inequalities

$$\begin{aligned}
& w^{\star\prime}_{\underset{\sim}{\lambda}}\mu - \frac{\gamma}{2} w^{\star\prime}_{\underset{\sim}{\lambda}}\Sigma w^{\star}_{\underset{\sim}{\lambda}} - \underset{\sim}{\lambda}||w^{\star}_{\underset{\sim}{\lambda}}||_q \\
\geq \; & w^{\star\prime}_{\tilde{\lambda}}\mu - \frac{\gamma}{2} w^{\star\prime}_{\tilde{\lambda}}\Sigma w^{\star}_{\tilde{\lambda}} - \underset{\sim}{\lambda}||w^{\star}_{\tilde{\lambda}}||_q \\
= \; & w^{\star\prime}_{\tilde{\lambda}}\mu - \frac{\gamma}{2} w^{\star\prime}_{\tilde{\lambda}}\Sigma w^{\star}_{\tilde{\lambda}} - \tilde{\lambda}||w^{\star}_{\tilde{\lambda}}||_q - (\underset{\sim}{\lambda} - \tilde{\lambda})||w^{\star}_{\tilde{\lambda}}||_q \\
\geq \; & w^{\star\prime}_{\underset{\sim}{\lambda}}\mu - \frac{\gamma}{2} w^{\star\prime}_{\underset{\sim}{\lambda}}\Sigma w^{\star}_{\underset{\sim}{\lambda}} - \tilde{\lambda}||w^{\star}_{\underset{\sim}{\lambda}}||_q - (\underset{\sim}{\lambda} - \tilde{\lambda})||w^{\star}_{\tilde{\lambda}}||_q \\
= \; & w^{\star\prime}_{\underset{\sim}{\lambda}}\mu - \frac{\gamma}{2} w^{\star\prime}_{\underset{\sim}{\lambda}}\Sigma w^{\star}_{\underset{\sim}{\lambda}} - \underset{\sim}{\lambda}||w^{\star}_{\underset{\sim}{\lambda}}||_q - (\underset{\sim}{\lambda} - \tilde{\lambda})\left(||w^{\star}_{\tilde{\lambda}}||_q - ||w^{\star}_{\underset{\sim}{\lambda}}||_q\right)
\end{aligned}$$

The inequalities imply that

$$(\underset{\sim}{\lambda} - \tilde{\lambda})\left(||w^{\star}_{\tilde{\lambda}}||_q - ||w^{\star}_{\underset{\sim}{\lambda}}||_q\right) \geq 0$$

Hence for $\underset{\sim}{\lambda} \leq \tilde{\lambda}$ it has to follow that $||w^{\star}_{\underset{\sim}{\lambda}}||_q \geq ||w^{\star}_{\tilde{\lambda}}||_q$. □

4.A.2 Proof to Proposition 4.2.2

Proof. As $\lambda \to \infty$ the first optimization problem reduces to

$$\min_{w:w'\iota=1} ||w||_q \tag{4.12}$$

Let $q = 1$: The constraint is that the weights sum up to one, i.e. $1 = w'\iota = \sum_i w_i = \sum_{i:w_i\geq 0}|w_i| - \sum_{i:w_i<0}|w_i|$. Denote the short position by $\delta = \sum_{i:w_i<0}|w_i|$. Then $\sum_{i:w_i\geq 0}|w_i| = 1+\delta$. The norm of the weights are given by $||w||_1 = \sum_{i:w_i\geq 0}|w_i| + \sum_{i:w_i<0}|w_i| = 1+\delta+\delta = 1+2\delta$. Hence the minimum norm is given by $||w||_1 = 1$ iff the portfolio has no short positions. The constrained optimization of eq. 4.2 is then equivalent to

$$\max_{w:w'\iota=1,w_i\geq 0} w'\mu - \frac{\gamma}{2}w'\Sigma w$$

which is exactly the no short-sale portfolio.

Let $q = 2$: It is easy to see that problem 4.12 is minimized for $w_i = \frac{1}{n}$. □

4.A.3 Proof to Proposition 4.2.3

Proof. We prove part a.) first. As the returns are not too strongly correlated, we analyze the case when returns are non-positively correlated. Then the covariance matrix can be split up into

$$\Sigma = sI - B$$

with $B \geq 0$ (element-wise), $\rho(B) \leq s \in \mathbb{R}^+$ with ρ being the spectral radius of B. Parameter s can be selected such that the conditions are fulfilled. For large s, $\frac{B}{s} \to I$ with $\rho(B) = s$. Then Σ is a $M-$matrix. The inverse of a $M-$matrix is non-negative (Berman and Plemmons, 1994), i.e. $\Sigma^{-1} \geq 0$ (element-wise). The term $\Sigma^{-1}\mu \geq 0$ as $\mu \geq 0$ and $\Sigma^{-1}\iota \geq 0$ by definition. Hence for the norms it holds that $||\Sigma^{-1}\mu||_1 = \iota'\Sigma^{-1}\mu$ and $||\Sigma^{-1}\iota||_1 = \iota'\Sigma^{-1}\iota$. As the closed form solution of $w^\star$ is given by (Best and Grauer, 1991a)

$$w^\star = \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\mu$$

Then the L_1 norm of the weights is given by

$$\begin{aligned}
||w^\star||_1 &= ||\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\mu||_1 \\
&\leq ||\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}||_1 + \frac{1}{\gamma\iota'\Sigma^{-1}\iota}||(\iota'\Sigma^{-1}\iota)\Sigma^{-1}\mu - \Sigma^{-1}\iota\iota'\Sigma^{-1}\mu||_1 \\
&= \frac{\iota'\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{1}{\gamma\iota'\Sigma^{-1}\iota}\left(\iota'\Sigma^{-1}\iota\iota'\Sigma^{-1}\mu - \iota'\Sigma^{-1}\iota\iota'\Sigma^{-1}\mu\right) \\
&= 1 + 0 = 1
\end{aligned}$$

Note that the restriction that Σ is a $M-$matrix is sufficient but not necessary that $||w^\star||_1 = 1$. For slight changes of Σ it still holds that $\Sigma^{-1}\mu \geq 0$ and $\Sigma^{-1}\iota \geq 0$. In this case it holds that $||w^\star||_1 = 1$. We conclude that for moderate levels of (positive) correlation in the covariance matrix the L_1 norm of the optimal weights is small.

Part b.) first assumes that $\mu = \underline{\mu}\iota$ with $\underline{\mu} \in \mathbb{R}$. We find that

$$\begin{aligned} w^\star &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{\underline{\mu}}{\gamma}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}}{\iota'\Sigma^{-1}\iota}\right)\iota \\ &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{\underline{\mu}}{\gamma}\left(\Sigma^{-1}\iota - \frac{\Sigma^{-1}\iota\iota'\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}\right) \\ &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} + \frac{\underline{\mu}}{\gamma}\left(\Sigma^{-1}\iota - \Sigma^{-1}\iota\right) \\ &= \frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} \end{aligned}$$

As the portfolios are only positively correlated due to a common risk factor we can write

$$\Sigma = \sigma\left(\iota\iota' + \Gamma\right)$$

where $\sigma\Gamma$ is the covariance matrix of the idiosyncratic risk and $\sigma\iota\iota'$ represents the systemic risk. By assumption the positive correlation stems from the common risk factor $\sigma\iota\iota'$. Similar to part a.) we can split up $\Gamma = sI - B$ and show that $\Gamma^{-1} \geq 0$. Going back to Σ we can apply the Sherman-Morrison formula (Sherman and Morrison, 1949) and show that

$$\Sigma^{-1} = \frac{1}{\sigma}\left[\Gamma^{-1} - \frac{\Gamma^{-1}\iota\iota'\Gamma^{-1}}{1 + \iota'\Gamma^{-1}\iota}\right]$$

Using $\iota'\Gamma^{-1}\iota \geq 0$ we find that

$$\begin{aligned} \Sigma^{-1}\iota &= \frac{1}{\sigma}\left[\Gamma^{-1}\iota - \frac{\Gamma^{-1}\iota\iota'\Gamma^{-1}\iota}{1 + \iota'\Gamma^{-1}\iota}\right] \\ &= \frac{1}{\sigma}\left[\frac{\Gamma^{-1}\iota}{1 + \iota'\Gamma^{-1}\iota}\right] \geq 0 \end{aligned}$$

With the same argumentation as in part a.) for the norm of the weight it holds

$$\begin{aligned} \|w^\star\|_1 &= \left\|\frac{\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota}\right\|_1 \\ &= \frac{\iota'\Sigma^{-1}\iota}{\iota'\Sigma^{-1}\iota} = 1 \end{aligned}$$

□

4.A.4 Proof to Proposition 4.2.4

Proof. The (estimated) objective function is given by CE (and $\hat{CE}$, resp.). The optimal weights are given by $w^\star$ and the estimated counterparts by $\hat{w}$. As the weights are optimized with respect to the given objective function it holds that $\hat{CE}(w) - \hat{CE}(\hat{w}) \leq 0$ for all weights w. Then we find the following relationship

$$
\begin{aligned}
0 &\leq CE(w^\star) - CE(\hat{w}) \\
&= CE(w^\star) - \hat{CE}(w^\star) + \underbrace{\hat{CE}(w^\star) - \hat{CE}(\hat{w})}_{\leq 0} + \hat{CE}(\hat{w}) - CE(\hat{w}) \\
&\leq CE(w^\star) - \hat{CE}(w^\star) + \hat{CE}(\hat{w}) - CE(\hat{w}) \\
&= CE_\lambda(w^\star) - \hat{CE}_\lambda(w^\star) + \hat{CE}_\lambda(\hat{w}) - CE_\lambda(\hat{w}) \\
&\leq 2 \cdot \max_{w \in \{w^\star, \hat{w}\}} |CE_\lambda(w) - \hat{CE}_\lambda(w)|
\end{aligned}
$$

□

4.A.5 Proof to Proposition 4.2.5

Proof. The part of the proposition is easily calculated by

$$
\begin{aligned}
|CE_\lambda(w) - \hat{CE}_\lambda(w)| &= |CE_\lambda(w) - \hat{CE}_\lambda(w)| \\
&= |w'(\mu - \hat{\mu}) - \frac{\gamma}{2} w' \left(\Sigma - \hat{\Sigma}\right) w| \\
&\leq ||\mu - \hat{\mu}||_\infty ||w||_1 + \frac{\gamma}{2} ||\Sigma - \hat{\Sigma}||_\infty ||w||_1^2
\end{aligned}
$$

The second part is a result of Theorem 2 in Fan et al. (2012). For the convergence rate of the error in the sample mean see the proof of the cited theorem. □

Appendix 4.B Algorithm

4.B.1 Iterative algorithm

We use the algorithm of Loris and Verhoeven (2012). The algorithm optimizes

$$\hat{x} = \arg\min_x ||Kx - y||_2^2 + 2\lambda ||Ax||_1 \quad s.t.\ Bx = b$$

They show that the optimal solution can be obtained by the following iterative algorithm.

$$\begin{aligned}
\overline{v}_{n+1} &= v_n - (Bx_n - b) \\
\overline{x}_{n+1} &= x_n + \tau_1 K'(y - Kx_n) + \tau_3 B'\overline{v}_{n+1} - \tau_1 A' w_n \\
w_{n+1} &= \mathbb{P}_\lambda \left(w_n + \frac{\tau_2}{\tau_1} A\overline{x}_{n+1} \right) \\
x_{n+1} &= x_n + \tau_1 K'(y - Kx_n) + \tau_3 B'\overline{v}_{n+1} - \tau_1 A' w_{n+1} \\
v_{n+1} &= v_n - \frac{1}{\alpha}(Bx_{n+1} - b)
\end{aligned}$$

with step size parameters $\tau_1, \tau_2, \tau_3 > 0$ such that $||\tau_1 K'K/2 + \tau_3 B'B|| < 1$ and $\tau_2||AA'|| < 1$. The projection on the l_∞ ball of radius λ is defined by

$$P_\lambda(z) = \begin{cases} \frac{z}{|z|}\lambda, & |z| > \lambda \\ z, & |z| \leq \lambda \end{cases}$$

Loris and Verhoeven (2012) show that algorithm converges for $\alpha > \frac{1}{2}$ to the true value, e.g. $x_n \overset{n\to\infty}{\longrightarrow} \hat{x}$. We set $\tau_1 = 1/(||K'K|| + 1)$, $\tau_2 = 1/(||AA'|| + 1)$, $\tau_3 = 1/(2||B'B|| + 1)$, $\alpha = \frac{3}{4}$.

4.B.2 Relationship to CE

The iterative algorithm presented in Section 4.B.1 solves the minimization of a L_1-norm penalized least squares functional under additional linear constraints. In our context a CE has to be maximized under the constraint that weights sum to one. We show that our problem is a special case of the general penalized least squares problem.

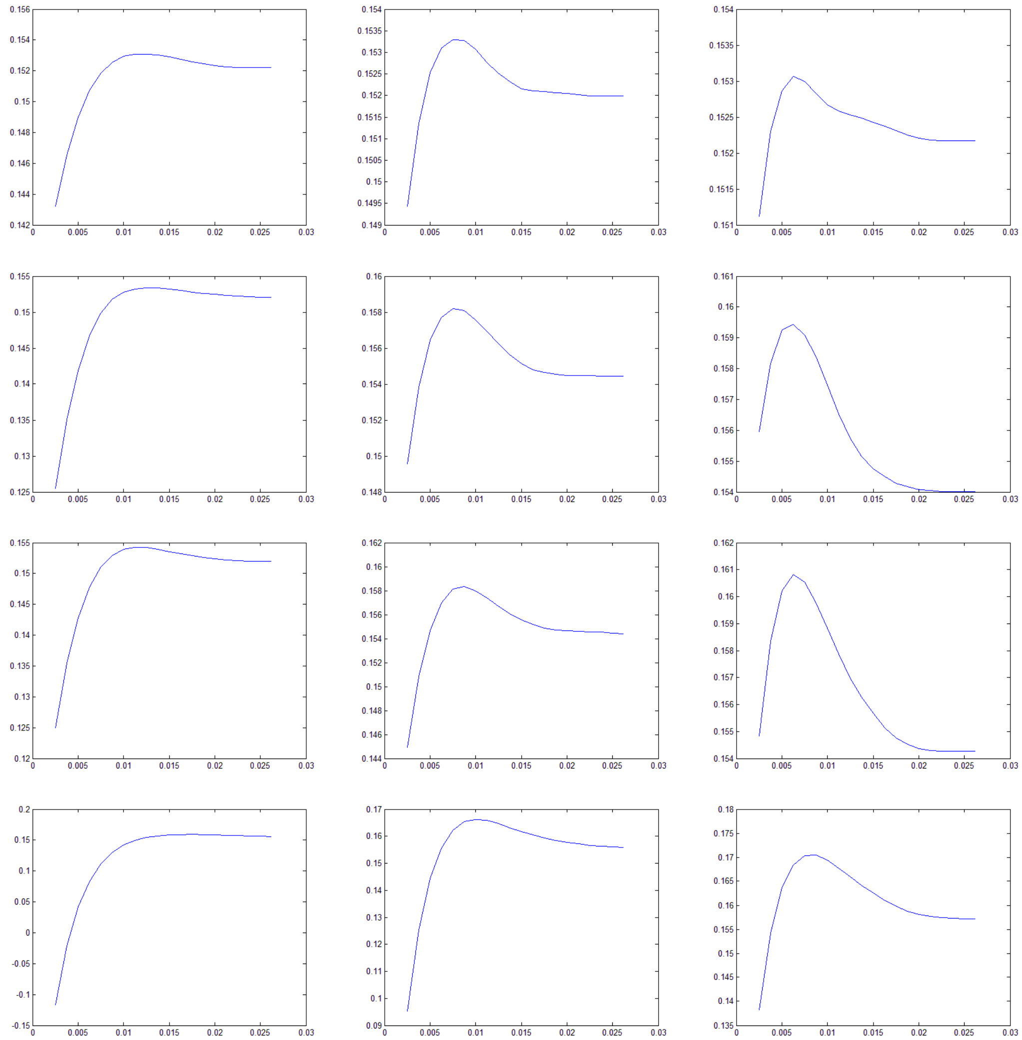
$$\begin{aligned}
& \arg\min_x ||Kx - y||_2^2 + 2\lambda||Ax||_1 \quad s.t.\ Bx = b \\
\iff\quad & \arg\min_{x:Bx=b} x'K'Kx - 2x'K'y + y'y + 2\lambda||Ax||_1 \\
\iff\quad & \arg\max_{x:Bx=b} x'2K'y - x'K'Kx - 2\lambda||Ax||_1 + const \\
\iff\quad & \arg\max_{x:Bx=b} x'\mu - x'\frac{\gamma}{2}\Sigma x - 2\lambda||Ax||_1 \\
\implies\quad & \arg\max_{x:\iota'x=1} CE(x) - \tilde{\lambda}||x||_1
\end{aligned}$$

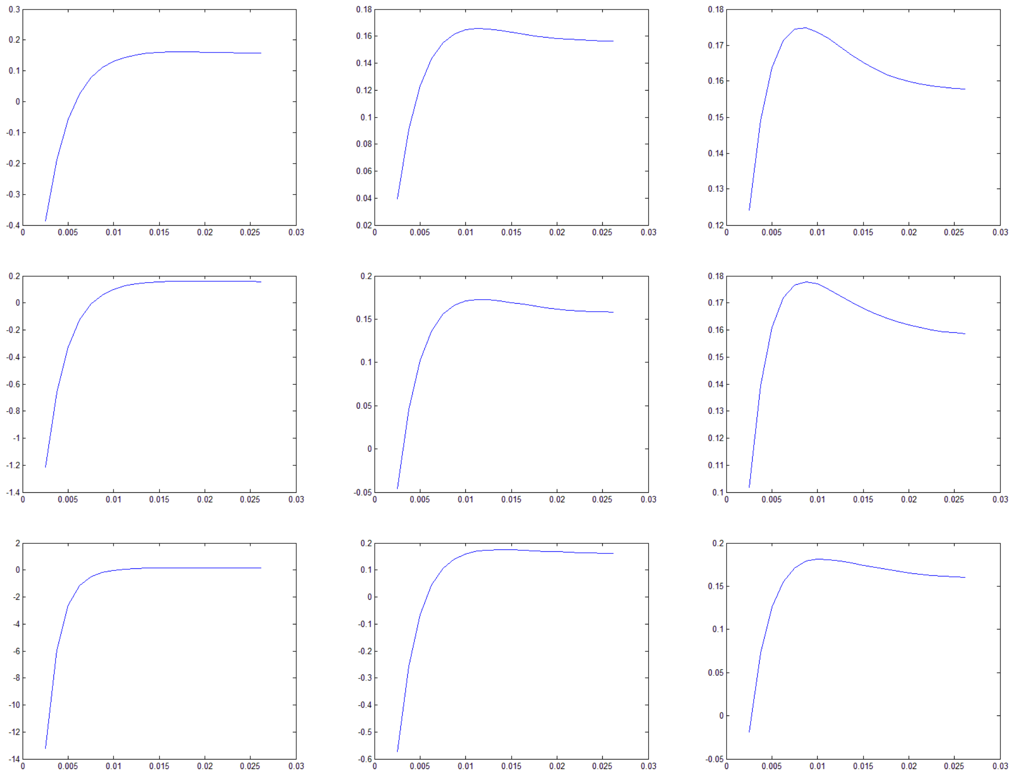
We find that both optimization problems are equivalent. The parameter of interests are given by $B = \iota'$, $b = 1$, $A = I_n$, $y = \frac{1}{2}K'^{-1}\mu$, $\tilde{\lambda} = 2\lambda$ and K being the Cholesky decomposition of $\frac{\gamma}{2}\Sigma$, e.g. $K = chol(\frac{\gamma}{2}\Sigma)$.

Appendix 4.C Data and Results

4.C.1 Simulation

Based on the procedure described in Section 4.3.1 the CEs are estimated for different levels of $\lambda \in [0, 0.025]$. As common number of assets, we used $n \in \{6, 10, 12, 25, 30, 36, 48\}$ (top to bottom) and length of estimation window $T \in \{60, 90, 120\}$ (left to right).

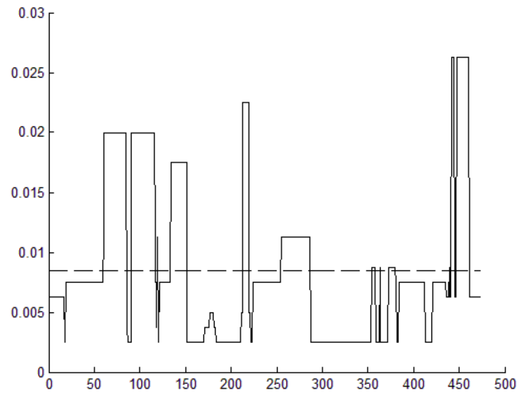




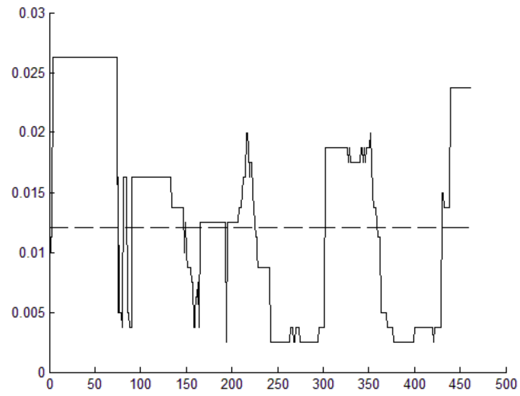
4.C.2 Optimal Penalty

For different data sets, we calculate the optimal penalty level $\hat{\lambda}$ based on the past T observations. The weights $\hat{w}_{\lambda}$ which obtain the best out-of-sample CE in the last T month, determine the best estimated penalty $\hat{\lambda}$. The optimal penalty level appears to be unstable and difficult to forecast. We find no optimal penalty level for the Ind38 portfolio. The moments of the return are estimated on T observations, the performance of the model is estimated on T observations, and the optimal penalty level needs additional T observations. As the available return period for the Ind38 portfolio is shorter than $3 \cdot T$, an optimal penalty level could not be calculated in this case.

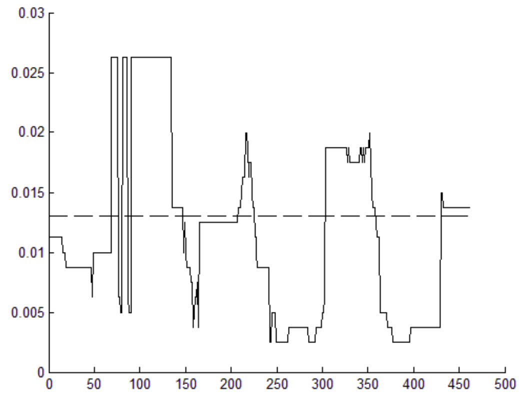
The plots refer to the data sets Ind5/Ind10, Ind12/Ind17, Ind30/Ind48 from top left to bottom right. The dashed line denotes the penalty level $0.3 \log n/T$.



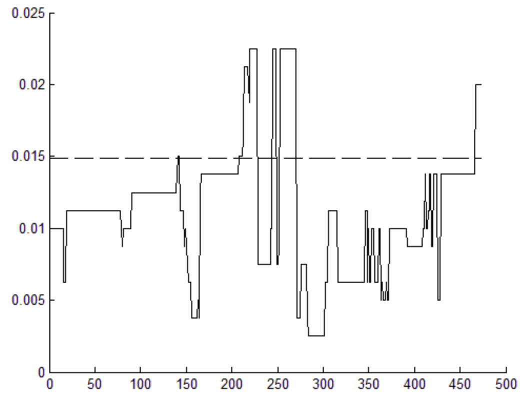
Ind5



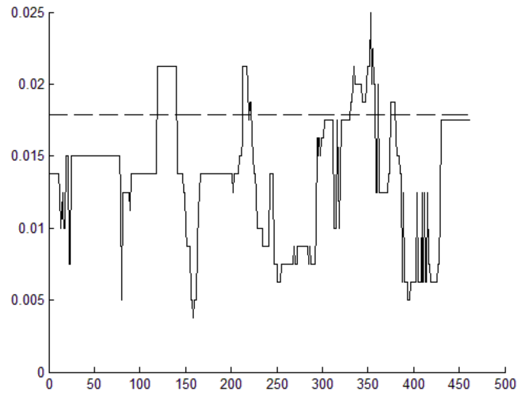
Ind10



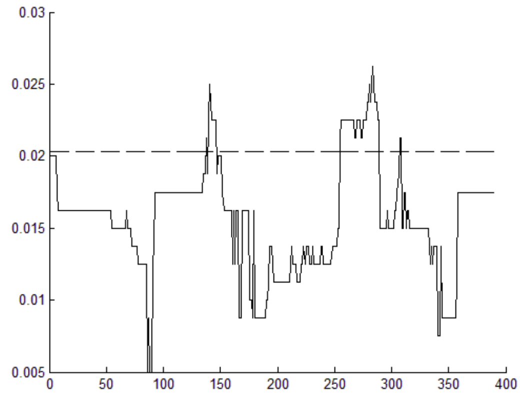
Ind12



Ind17



Ind30

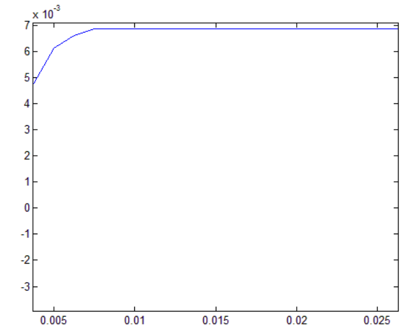


Ind48

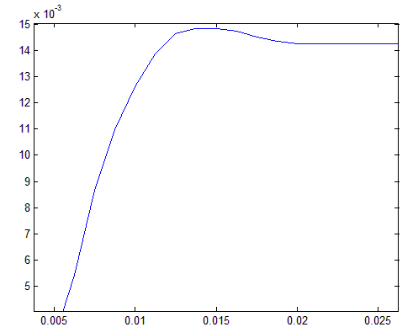
4.C.3 Empirical Study

We calculate the performance of weights optimized according to eq. 4.3. The penalty λ is given on the $x-$axis, while the realized CE is stated on the $y-$axis.
The plots refer to the data sets Ind5/Ind10, Ind12/Ind17, Ind30/Ind38 and Ind48

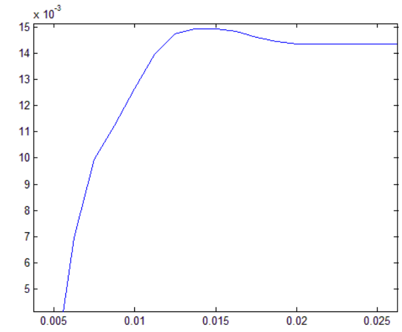
from top left to bottom right.



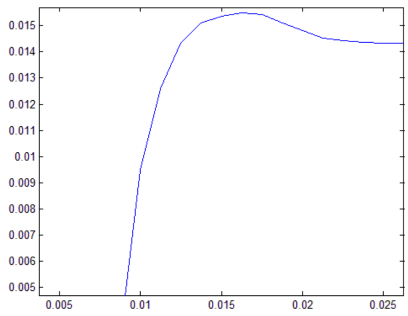
Ind5



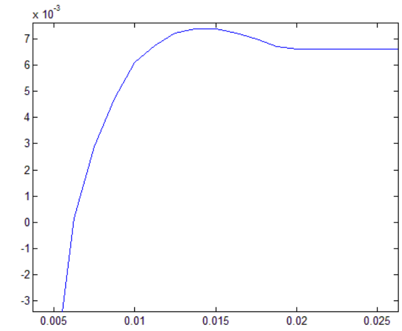
Ind10



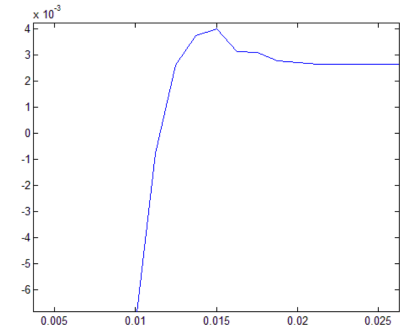
Ind12



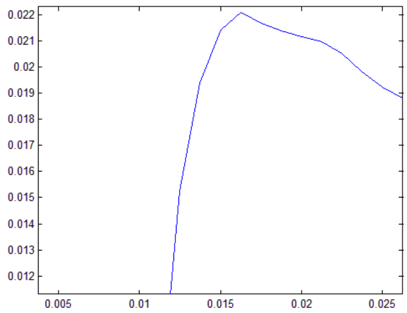
Ind17



Ind30



Ind38



Ind48

Appendix 4.D Robustness Checks

4.D.1 Robustness of Risk Aversion

In the simulation study, we used a risk aversion parameter of $\gamma = 2$. As a robustness check we also consider risk aversion of $\gamma = 0.5$ and $\gamma = 5$.

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0004	0.0002	1.9858	0.0480
log(n)/T	0.309756	0.001199	258.4161	0
R-squared	0.947797	Mean dependent var		0.011466
Adjusted R-squared	0.947797	S.D. dependent var		0.003516
S.E. of regression	0.000803	Akaike info criterion		-11.4125
Sum squared resid	0.000194	Schwarz criterion		-11.40019
Log likelihood	1718.582	Durbin-Watson stat		1.114017

Table 4.D.4: Regression of $\lambda^\star$ on $\log(n)/T$ with a constant and $\gamma = 0.5$.

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
log(n)/T	0.309507	0.004201	73.67951	0
R-squared	0.947797	Mean dependent var		0.011466
Adjusted R-squared	0.947623	S.D. dependent var		0.003516
S.E. of regression	0.000805	Akaike info criterion		-11.40587
Sum squared resid	0.000194	Schwarz criterion		-11.38124
Log likelihood	1718.584	F-statistic		5428.67
Durbin-Watson stat	1.114299	Prob(F-statistic)		0

Table 4.D.5: Regression of $\lambda^\star$ on $\log(n)/T$ without a constant and $\gamma = 0.5$.

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000973	0.000343	2.834139	0.0049
log(n)/T	0.308624	0.008888	34.72253	0
R-squared	0.801283	Mean dependent var		0.012396
Adjusted R-squared	0.800619	S.D. dependent var		0.003813
S.E. of regression	0.001702	Akaike info criterion		-9.906914
Sum squared resid	0.000867	Schwarz criterion		-9.882282
Log likelihood	1492.991	F-statistic		1205.654
Durbin-Watson stat	0.969708	Prob(F-statistic)		0

Table 4.D.6: Regression of $\lambda^\star$ on $\log(n)/T$ with a constant and $\gamma = 5$.

Dependent Variable: λ				
Method: Least Squares				
Included observations: 301				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
log(n)/T	0.332763	0.00257	129.4753	0
R-squared	0.795945	Mean dependent var		0.012396
Adjusted R-squared	0.795945	S.D. dependent var		0.003813
S.E. of regression	0.001722	Akaike info criterion		-9.887049
Sum squared resid	0.00089	Schwarz criterion		-9.874733
Log likelihood	1489.001	Durbin-Watson stat		0.966864

Table 4.D.7: Regression of $\lambda^\star$ on $\log(n)/T$ without a constant and $\gamma = 5$.

Bibliography

BERMAN, A. AND R. PLEMMONS (1994): *Nonnegative Matrices in the Mathematical Sciences*, SIAM.

BEST, M. AND R. GRAUER (1991a): "On the sensitivity of mean variance efficient portfolios to changes in asset means: some analytical and computational results," *Review of Financial Studies*, 4 (2), 315–342.

——— (1991b): "Sensitivity Analysis for Mean-Variance Portfolio Problems," *Management Science*, 37 (8), 980–989.

BRODIE, J., I. DAUBECHIES, C. DE MOL, D. GIANNONE, AND I. LORIS (2009): "Sparse and stable Markowitz portfolios," *PNAS*, 106(30), 12267–12272.

CHAN, L. K. C., J. KARCESKI, AND J. LAKONISHOK (1999): "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model," *Review of Financial Studies*, 12, 937–974.

DEMIGUEL, V., L. GARLAPPI, F. J. NOGALES, AND R. UPPAL (2009a): "Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms." *Management Science*, 55(5), 798–812.

DEMIGUEL, V., L. GARLAPPI, AND R. UPPAL (2009b): "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22 (5), 1915–1953.

EFRON, B. AND G. GONG (1983): "A leisurely look at the bootstrap, the jackknife, and cross-validation." *Amer. Statistician*, 1, 36–48.

EFRON, B., T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI (2004): "Least angle regression." *The Annals of Statistics*, 32(2), 407–499.

FAN, J., J. ZHANG, AND K. YU (2012): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios," *Journal of American Statistical Association*, 107, 592–606.

FROST, P. A. AND J. E. SAVARINO (1986): "An empirical Bayes approach to portfolio selection." *Journal of Financial and Quantitative Analysis*, 21, 293–305.

JAGANNATHAN, R. AND T. MA (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps." *Journal of Finance*, 54(4), 1651–1684.

JORION, P. (1986): "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21(3), 2279–292.

——— (1991): "Bayesian and CAPM estimators of the means: Implications for portfolio selection." *Journal of Banking and Finance*, 15, 717–727.

KEMPF, A. AND C. MEMMEL (2006): "Estimating the global minimum variance portfolio," *Schmalenbach Business Review*, 58, 332–348.

LEDOIT, O. AND M. WOLF (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 21(1), 110–119.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2011): "Robust performance hypothesis testing with the variance," *Wilmott Magazine*, 86–89.

LORIS, I. AND C. VERHOEVEN (2012): "An iterative algorithm for sparse and constrained recovery with applications to divergence-free current reconstructions in magneto-encephalography," *Computational Optimization and Applications*, 1–18.

MARKOWITZ, H. M. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

MERTON, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics*, 8, 323–361.

SCHANBACHER, P. (2012): "Averaging across Asset Allocation Models," *Working Paper*.

SHERMAN, J. AND W. J. MORRISON (1949): "Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix," *Annals of Mathematical Statistics*, 20, 620–624.

TIBSHIRANI, R. (1996): "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society Series B*, 58(1), 267–288.

Tychonoff, A. N. and V. Y. Arsenin (1977): *Solution of Ill-posed Problems*, Winston & Sons.

Complete Bibliography

Andrews, D. and J. Monahan (1992): "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953–966.

Arrow, K. (1971): *Aspects of the Theory of Risk Bearing*, Markham, chap. The theory of risk aversion, 90–109.

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9 (3), 203–228.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1997): "Thinking coherently," *RISK*, 10 (11), 68–71.

Bates, J. and C. Granger (1969): "The combination of forecasts," *Operational Research Quarterly*, 20, 451–468.

Belousova, J. and G. Dorfleitner (2012): "On the diversification benefits of commodities from the perspective of euro investors," *Journal of Banking and Finance*, 36, 2455–2472.

Berman, A. and R. Plemmons (1994): *Nonnegative Matrices in the Mathematical Sciences*, SIAM.

Best, M. and R. Grauer (1991a): "On the sensitivity of mean variance efficient portfolios to changes in asset means: some analytical and computational results," *Review of Financial Studies*, 4 (2), 315–342.

——— (1991b): "Sensitivity Analysis for Mean-Variance Portfolio Problems," *Management Science*, 37 (8), 980–989.

Biglova, A., S. Ortobelli, S. Rachev, and S. Stoyanov (2004): "Different Approaches to Risk Estimation in Portfolio Theory," *The Journal of Portfolio Management*, 31 (1), 103–112.

Bodie, Z., A. Kane, and A. J. Marcus (1993): *Investments*, Homewood.

Brainard, K. (2009): "Summary of Findings for FY 2008," *Public Fund Survey*, available from: www.publicfundsurvey.org, 10.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 78, 1–3.

Britten-Jones, M. (1999): “The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights.” *Journal of Finance*, 54(2), 655–671.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009): “Sparse and stable Markowitz portfolios,” *PNAS*, 106(30), 12267–12272.

Cai, X., K. Teo, X. Yang, and X. Zhou (2000): “Portfolio optimization under a minimax rule,” *Management Science*, 46 (7), 957–972.

Chan, L. K. C., J. Karceski, and J. Lakonishok (1999): “On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model,” *Review of Financial Studies*, 12, 937–974.

Chateauneuf, A., F. Maccheroni, M. Marinacci, and J.-M. Tallon (2005): “Monotone continuous multiple priors,” *Economic Theory*, 26 (4), 973–982.

Cho, D. (2011): “Estimation risk in covariance,” *Journal of Asset Management*, 12, 248–259.

Clemen, R. (1989): “Combining forecasts: A review and annotated bibliography,” *International Journal of Forecasting*, 5, 559–581.

Clements, M. P. and D. I. Harvey (2011): “Combining Probability Forecasts,” *International Journal of Forecasting*, 27, 208–223.

DeMiguel, V., L. Garlappi, F. J. Nogales, and R. Uppal (2009a): “Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms.” *Management Science*, 55(5), 798–812.

DeMiguel, V., L. Garlappi, and R. Uppal (2009b): “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?” *Review of Financial Studies*, 22 (5), 1915–1953.

Diebold, F. and P. Pauly (1987): “Structural change and the combination of forecasts,” *Journal of Forecasting*, 6, 21–40.

Diebold, F. X., T. A. Gunther, and A. Tay (1998): “Evaluating Density Forecasts, with Applications to Financial Risk Management,” *International Economic Review*, 39, 863–883.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

Ding, Y. (2006): "Portfolio Selection under Maximum Minimum Criterion," *Quality and Quantity*, 40 (3), 457–468.

Efron, B. and G. Gong (1983): "A leisurely look at the bootstrap, the jackknife, and cross-validation." *Amer. Statistician*, 1, 36–48.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004): "Least angle regression." *The Annals of Statistics*, 32(2), 407–499.

Elliott, G. (2011): "Averaging and the Optimal Combination of Forecasts," *Working Paper*.

Embrechts, P., A. J. McNeil, and D. Straumann (2002): *Correlation and Dependence in Risk Management: Properties and Pitfalls*, Cambridge University Press.

Fama, E. F. and K. R. French (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33 (1), 3–56.

Fan, J., J. Zhang, and K. Yu (2012): "Asset Allocation and Risk Assessment with Gross Exposure Constraints for Vast Portfolios," *Journal of American Statistical Association*, 107, 592–606.

Foellmer, H. and A. Schied (2002): "Convex measures of risk and trading constraints," *Finance and Stochastics*, 6 (4), 429–447.

——— (2010): *Encyclopedia of Quantitative Finance*, John Wiley & Sons, chap. Coherent and convex risk measures, 355–363.

Frahm, G. (2012): "Capital Allocation Under Risk and Uncertainty," *Working Paper*.

Frost, P. A. and J. E. Savarino (1986): "An empirical Bayes approach to portfolio selection." *Journal of Financial and Quantitative Analysis*, 21, 293–305.

Garlappi, L., R. Uppal, and T. Wang (2007): "Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach," *The Review of Financial Studies*, 20, 41–81.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013): "Combining

Expert Forecasts: Can Anything Beat the Simple Average?" *International Journal of Forecasting*.

Geweke, J. and G. Amisano (2011): "Optimal prediction pools," *Journal of Econometrics*, 164, 130–141.

Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18 (2), 141–153.

Gneiting, T., F. Balabdaoui, and A. Raftery (2007): "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society*, 69, 243–268.

Gneiting, T. and E. A. Raftery (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Gneiting, T. and R. Ranjan (2011): "Combining Predictive Distributions," *Working Paper*.

Good, I. J. (1952): "Rational Decisions," *Journal of the Royal Statistical Society*, 14, 107–114.

Greene, W. H. (2002): *Econometric Analysis*, Prentice-Hall.

Hendrickson, A. D. and R. J. Buehler (1971): "Proper Scores for Probability Forecasters," *The Annals of Mathematical Statistics*, 42, 1916–1921.

Hodder, J., J. Jackwerth, and O. Kolokolova (2012): "Improved Portfolio Choice Using Second-Order Stochastic Dominance," *Working Paper*.

Jagannathan, R. and T. Ma (2003): "Risk reduction in large portfolios: Why imposing the wrong constraints helps." *Journal of Finance*, 54(4), 1651–1684.

Jorion, P. (1986): "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21(3), 2279–292.

——— (1991): "Bayesian and CAPM estimators of the means: Implications for portfolio selection." *Journal of Banking and Finance*, 15, 717–727.

Kan, R. and G. Zhou (2007): "Optimal Portfolio Choice with Parameter Uncertainty," *Journal of Financial and Quantitative Analysis*, 42(3), 621–656.

Kascha, C. and F. Ravazzolo (2010): "Combining inflation density forecasts." *Journal of Forecasting*, 29, 231–250.

Kempf, A. and C. Memmel (2006): "Estimating the global minimum variance portfolio," *Schmalenbach Business Review*, 58, 332–348.

Krueger, F. and P. Schanbacher (2012): "Strictly Proper Scoring Rules and the Wisdom of the Crowd," *University of Konstanz, Working Paper.*

Larrick, R. P. and J. B. Soll (2006): "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle," *Management Science*, 52, 111–127.

Ledoit, O. and M. Wolf (2003): "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." *Journal of Empirical Finance*, 10(5), 603–621.

——— (2004a): "Honey, I Shrunk the Sample Covariance Matrix," *Journal of Portfolio Management*, 21(1), 110–119.

——— (2004b): "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, 88, 365–411.

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2011): "Robust performance hypothesis testing with the variance," *Wilmott Magazine*, 86–89.

Lichtendahl, K. C., Y. Grushka-Cockayne, and R. L. Winkler (2012): "Is it Better to Average Probabilities or Quantiles?" *Management Science (forthcoming).*

Lintner, J. and W. Krasker (1982): "Minimax behavior in portfolio selection," *Journal of Finance*, 37 (2), 609–614.

Loris, I. and C. Verhoeven (2012): "An iterative algorithm for sparse and constrained recovery with applications to divergence-free current reconstructions in magneto-encephalography," *Computational Optimization and Applications*, 1–18.

Mansini, R., W. Ogryczak, and G. Speranza (2003): "LP solvable models for portfolio optimization: a classification and computational comparison," *IMA Journal of Management Mathematics*, 14 (3), 187–220.

Manski (2010): "When Consensus Choice Dominates Individualism: Jensen's Inequality and Collective Decisions Under Uncertainty," *Quantitative Economics*, 1, 187–202.

Markowitz, H. M. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

McNees, S. K. (1992): "The Uses and Abuses of "Consensus" Forecasts," *Journal of Forecasting*, 11, 703–710.

Merton, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation." *Journal of Financial Economics*, 8, 323–361.

Michaud, R. (1989): "The Markowitz optimization enigma: Is optimized optimal?" *Financial Analysts Journal*, 45, 31–42.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Politis, D. and J. Romano (1994): "The stationary bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313.

Pratt, J. (1964): "Risk aversion in the small and in the large," *Econometrica*, 32, 122–136.

Press, J. (1972): *Applied Multivariate Analysis*, Holt, Rinehart and Winston, New York.

R Development Core Team (2012): "R: A Language and Environment for Statistical Computing," *http://www.R-project.org*.

Radcliff, R. C. (1990): *Investment Concepts, Analysis, Strategy*, HarperCollins.

Ranjan, R. and T. Gneiting (2010): "Combining Probability Forecasts," *Journal of the Royal Statistical Society*, 72, 71–91.

Rudebusch, G. D. and J. C. Williams (2009): "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve," *Journal of Business and Economic Statistics*, 27 (4), 492–503.

Schanbacher, P. (2012): “Averaging across Asset Allocation Models,” *Working Paper.*

Schervish, M. J. (1989): “A General Method for Comparing Probability Assessors,” *Annals of Statistics*, 4, 1856–1879.

Sengupta, J. (1982): “A minimax policy for optimal portfolio choice,” *International Journal of Systems Science*, 13 (1), 39–56.

Sharpe, W. F. (1963): “A simplifed model for portfolio analysis.” *Management Science*, 9, 277–293.

——— (1964): “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” *The Journal of Finance*, 19(3), 425–442.

——— (1994): “The Sharpe Ratio,” *The Journal of Portfolio Management*, 21 (1), 49–58.

Sherman, J. and W. J. Morrison (1949): “Adjustment of an Inverse Matrix Corresponding to Changes in the Elements of a Given Column or a Given Row of the Original Matrix,” *Annals of Mathematical Statistics*, 20, 620–624.

Shuford, E., A. Albert, and H. E. Massengill (1966): “Admissible probability measurement procedures,” *Psychometrika*, 31, 125–145.

Smith, J. and K. Wallis (2009): “A simple explanation of the forecast combination puzzle,” *Oxford Bulletin of Economics and Statistics*, 71 (3), 331–355.

Staël von Holstein, C. A. S. (1970): *Assessment and Evaluation of Subjective Probability Distributions*, Economic Research Institute, Stockholm School of Economics; Norstedt (distr.).

Stock, J. and M. Watson (2004): “Combination forecasts of output growth in a seven-country data set,” *Journal of Forecasting*, 23, 405–430.

Stone, M. (1961): “The Opinion Pool,” *The Annals of Mathematical Statistics*, 32, 1339–1342.

Surowiecki, J. (2004): *The Wisdom of Crowds.*, RandomHouse, NewYork.

Teo, K. and X. Yang (2001): “Portfolio Selection Problem with Minimax Type Risk Function,” *Annals of Operations Research*, 101 (1), 333–349.

Tibshirani, R. (1996): “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society Series B*, 58(1), 267–288.

Toda, M. (1963): “Measurement of Subjective Probability Distributions,” *Division of Mathematical Psychology, Institute for Research, State College, Pennsylvania*, Report No. 3.

Tu, J. and G. Zhou (2011): “Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies,” *Journal of Financial Economics*, 99, 204–215.

Tuetuencue, R. and M. Koenig (2004): “Robust Asset Allocation,” *Annals of Operations Research*, 132, 157–187.

Tychonoff, A. N. and V. Y. Arsenin (1977): *Solution of Ill-posed Problems*, Winston & Sons.

Varian, H. R. (1992): *Microeconomic Analysis (Third ed.).*, W.W. Norton and Company.

Wallis, K. (2005): “Combining Density and Interval Forecasts: A Modest Proposal,” *Oxford Bulletin of Economics and Statistics*, 67, 983–994.

Weber, M., H. Jacobs, and S. Müller (2009): “How Should Private Investors Diversify? - An Empirical Evaluation of Alternative Asset Allocation Policies to Construct A World Market Portfolio,” *Working Paper*.

Young, M. (1998): “A Minimax Portfolio Selection Rule with Linear Programming Solution,” *Management Science*, 44 (5), 673–683.

Eigenabgrenzung

Kapitel 1 entstammt einer gemeinsamen Arbeit mit Fabian Krüger (University of Konstanz). Meine individuelle Leistung bei der Erstellung dieser Arbeit beträgt 50%.

Kapitel 2 wurde ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt.

Kapitel 3 entstammt einer gemeinsamen Arbeit mit Steffen Schaarschmidt (University of Konstanz). Meine individuelle Leistung bei der Erstellung dieser Arbeit beträgt 50%.

Kapitel 4 entstammt einer gemeinsamen Arbeit mit Prof. Dr. Winfried Pohlmeier (University of Konstanz). Meine individuelle Leistung bei der Erstellung dieser Arbeit beträgt 50%.